Australian Government
Department of Defence
Defence Science and
Technology Organisation

Construction of 3-D Audio Systems: Background, Research and General Requirements

Simon P.A. Parker, Geoffery Eberle, Russell L. Martin, and Ken I. McAnally

Air Operations Division
Defence Science and Technology Organisation

DSTO-TR-2184

ABSTRACT

Over the last few years one of the most promising advances for Human Machine Interfaces (HMI) has been the development of 3-Dimensional Audio (3-D Audio). The Air Operations Division of DSTO has been engaged in an extensive research program developing 3-D audio for the military aviation environment. This document is intended to provide some general background and to list some of the broader requirements that need to be considered when designing any 3-D audio system. Included in this report is some background information describing the perceptual basis of 3-D audio, a brief review of some of the research associated with the development of 3-D audio, a discussion of the hardware and software requirements for the implementation of 3-D audio, a brief survey of current commercial 3-D audio systems, and some discussion of the issues that require further consideration.

Published by

Air Operations Division
DSTO Defence Science and Technology Organisation
506 Lorimer St
Fishermans Bend, Victoria 3207 Australia

Telephone: (03) 9626 7000
Fax: (03) 9626 7999

AR-014-278
October 2008

APPROVED FOR PUBLIC RELEASE

Construction of 3-D Audio Systems: Background, Research and General Requirements

Executive Summary

Over the last few years one of the most promising advances for Human Machine Interfaces (HMI) has been the development of 3-Dimensional Audio (3-D Audio). The Air Operations Division of DSTO has been engaged in an extensive research program developing 3-D audio for the military aviation environment. Development has reached the point where a 3-D audio capability has been demonstrated in the laboratory that delivers equivalent performance between the virtual and the real world. Further evaluation is being conducted in simulation environments with similar results. The next challenge in the development of 3-D audio is to begin the process of transferring the technology from a laboratory environment to defence platforms in the field.

This report has been generated to provide an overview of relevant material associated with the construction of 3-D audio systems. Included in this report is some background information describing the perceptual basis of 3-D audio, a brief review of some of the research associated with the development of 3-D audio, a discussion of the hardware and software requirements for the implementation of 3-D audio, a brief survey of current commercial 3-D audio systems, and some discussion of the issues that require further consideration. It is not intended that this report be a detailed design guide for the construction of a 3-D audio system, since too many of the detailed design decisions are dependent on the specific application and the operational environment (i.e. platform and cockpit). Rather, this document is intended to provide some general background and to list some of the broader requirements that need to be considered when designing any 3-D audio system.

Authors

Simon Parker
Air Operations Division

Simon Parker is a Senior Research Scientist working in the Air Operations Division of DSTO as a human factors specialist. He has a BA (Hons) and a PhD in Psychology and investigated three-dimensional sound localisation in his doctoral work. Simon has interest and expertise in human factors issues associated with electronic warfare equipment, the design of auditory warnings, the design and synthesis of 3D Audio displays and the use of simulation in human factors research.

Geoffrey Eberle
Air Operations Division

Geoffrey Eberle is a Human Factors researcher within the Air Operations Division (AOD) of DSTO. He joined DSTO in 2000 after completing an honours degree in psychology. He has worked on a range of projects including 3-dimensional audio perception and virtual audio technologies, an area in which he has recently completed a Ph.D. His other area of interest is human-system performance modelling.

Russell Martin
Air Operations Division

Russell Martin is a Senior Research Scientist in Human Factors in Air Operations Division. He received a Ph.D. in Psychology from Monash University in 1988 and worked at the University of Queensland, Oxford University, the University of Melbourne and Deakin University prior to joining DSTO in 1995.

Ken McAnally
Air Operations Division

Ken McAnally is a Senior Research Scientist in Human Factors in Air Operations Division. He received a Ph.D. in Physiology and Pharmacology from the University of Queensland in 1990 and worked at the University of Melbourne, the University of Bordeaux and Oxford University prior to joining DSTO in 1996.

Contents

Acronyms and Abbreviations

3-D Audio	3 Dimensional Audio
A/D	Analog to Digital
ANR	Active Noise Reduction
AOD	Air Operations Division
AOSC	Air Operations Simulation Centre
AWACS	Airborne Warning and Control System
CEP	Communication Ear Plugs
D/A	Digital to Analog
DERA	Defence Research Agency
DSP	Digital Signal Processing
DSTO	Defence Science and Technology Organisation
DTF	Directional Transfer Function
EW	Electronic Warfare
HDD	Head Down Display
HMD	Helmet Mounted Display
HMI	Human Machine Interface
HRTF	Head Related Transfer Function
ICS	Internal Communication System
IID	Interaural Intensity Difference
ILD	Interaural Level Difference
ITD	Interaural Time Difference
SRT	Speech Reception Threshold
TCAS	Terrain Collision Avoidance Systems

1. Introduction

Over the last few years one of the most promising advances for Human Machine Interfaces (HMI) has been the development of 3-Dimensional Audio (3-D Audio). This type of audio places sounds outside the head at any location specified by the system. For example, when coupled with an Electronic Warfare system in an air platform, 3-D audio could be used to convey the location of an incoming missile by generating a sound that would appear to originate outside of the listener's head, at a location coincident with the direction of the incoming missile. One of the advantages of 3-D audio is that it makes use of a pre-existing ability that listeners use on an everyday basis. Each day listeners identify and track events that occur out of their field-of-view using auditory information. Because this ability is used everyday, restoring this capability to aircrew when in the cockpit allows them to identify and track critical events without large increases in workload.

The Air Operations Division of DSTO has been engaged in an extensive research program developing 3-D audio for the military aviation environment. Development has reached the point where a 3-D audio capability has been demonstrated in the laboratory that delivers equivalent performance between the virtual and the real world. Further evaluation is being conducted in simulation environments with similar results. The next challenge in the development of 3-D audio is to begin the process of transferring the technology from a laboratory environment to defence platforms in the field.

This report has been generated to provide an overview of relevant material associated with the construction of 3-D audio systems. Included in this report is some background information describing the perceptual basis of 3-D audio, a brief review of some of the research associated with the development of 3-D audio, a discussion of the hardware and software requirements for the implementation of 3-D audio, a brief survey of current commercial 3-D audio systems, and some discussion of the issues that require further consideration.

It is not intended that this report be a detailed design guide for the construction of a 3-D audio system, since too many of the detailed design decisions are dependent on the specific application and the operational environment (i.e. platform and cockpit). Rather, this document is intended to provide some general background and to list some of the broader requirements that need to be considered when designing any 3-D audio system. It should also be acknowledged that any review of this area will age rapidly as research into numerous aspects associated with 3-D audio continues, and as products associated with generating 3-D audio evolve in a rapidly developing market place.

2. Perceptual basis of 3-D Audio

As human beings we gather information about the state of the world around us in order to understand what is happening in our world and to help guide our actions. We use our sensory organs (i.e. our eyes, ears, etc) to gather this information, and the process of

interpreting the information we gather is called perception. Each sense provides a specific type of information, packaged in a form that is readily interpreted by the brain. 3-D audio is one type of perceptual information gathered by the auditory system to inform us about the location of events in our external world. This section details how information about the location of events is derived, and lists the various forms this information takes as it is provided to the brain. An understanding of the manner in which this information is provided and processed leads to an understanding of the requirements of any 3-D audio system which needs to replicate these different types of information if it is to create the appropriate perceptual experience.

The senses have evolved to work together to complement each other and provide a complete picture of what is happening around us. For example, the position of our eyes means that we have no visual information about what is happening behind us. We therefore use auditory information to gather information about what is occurring in those areas where we cannot see. A complete picture of the world around us therefore needs a balance of sensory information. This is not the case in aircraft cockpits where the auditory sense has not being fully utilised and where very little information is delivered to pilots via the auditory modality. This under-utilization is due in part to the way auditory information has been implemented in the cockpit, with the implementations bearing little resemblance to the way auditory information is presented on an everyday basis. Many examples of auditory warnings are artificial and quickly reach limits that make them unusable, failing to exploit the full potential of the auditory modality. This is true for our ability to determine object location on the basis of acoustic information alone that exists in our everyday world but has not been recreated in the cockpit. Our competence in this area has been well documented by research on sound localization over the last 100 years, and recent attempts to utilise this skill to provide location information about objects in the external environment have culminated in the production of what is now generally referred to as 3-D audio.

Our ability to derive the location of sounds arises from our use of different forms of information that arise from unique aspects of our physiology. The underlying design philosophy of 3-D audio is to recreate this information in exactly the same form as if a real sound were present in the external environment. The brain therefore responds in exactly the same manner and believes it hears a sound out in the environment at the appropriate location.

The information we use to determine the location of sounds can be considered as divided into two distinct types: binaural difference cues and spectral cues.

2.1 Binaural Difference Cues

Binaural difference cues arise from the fact that we have two ears. When a sound occurs on one side of the head, the presence of the head attenuates the sound wave. This causes a difference in perceived intensity between the two ears where the sound at the ear on the opposite side of the head will be perceived as being lower in intensity than the sound at the closest ear (see Figure 1). The extent to which this occurs is location dependent. For example, when a sound is located directly in front of the listener, the sound will be

perceived at the same intensity at both the left and right ears. This is also true for a sound source located directly behind the listener. As the sound moves away from these central positions, the intensity will increase in the ear closest to the sound source. The difference in intensity between the ears is referred to as the Interaural (i.e. inter-ear) Intensity Difference (IID, also known as Interaural Level Difference or ILD).

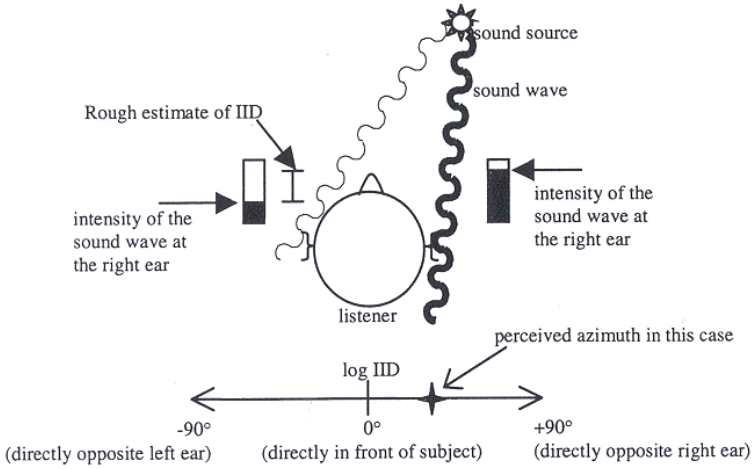


Figure 1: Change of Interaural Intensity Difference (IID) with location. Sound at the right ear is perceived at a higher intensity than the left ear because it is closer to the sound source (adapted from Cheng and Wakefield, 2001).

Just as the fact that we have two ears results in differences in intensity, having two ears means that there will be a difference in the time that a sound takes to travel to each ear (see Figure 2). For example, a sound located to the right of the listener will arrive earlier at the right ear than the left. As with interaural intensity differences, a sound source located directly in front or directly behind will have no difference in arrival time at the two ears. The difference in time of arrival (and phase) between the ears is referred to as the Interaural Time Difference (ITD).

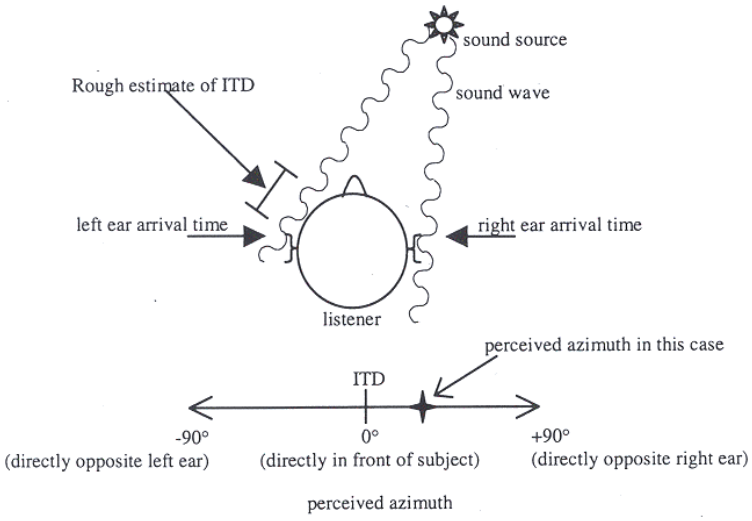


Figure 2: Change of Interaural Time Difference (ITD) with location. Sound at the right ear arrives earlier than at the left ear because it is closer to the sound source (adapted from Cheng and Wakefield, 2001).

2.2 Spectral Cues

Binaural difference cues alone are not enough however, to provide accurate localization of sounds. Each individual binaural difference can be associated with more than a single unique location. For example, if we imagine an axis drawn between the two ears that divides front from back (referred to as the interaural axis), then a sound 30 degrees in front of the interaural axis (Point "a" in Figure 3) will have the same pattern of binaural differences as a sound 30 degrees behind the interaural axis (Point "b" in Figure 3). This is also true for a sound 30 degrees above the interaural axis (Point "x" in Figure 3) and 30 degrees below (Point "y" in Figure 3). In fact if the head is assumed to be a sphere then the range of possible locations that are specified by any single binaural difference form a cone with the apex at the centre of the head (See Figure 3). This distribution of separate locations, each providing identical binaural differences, has come to be known as the 'cone of confusion'. If the only cues we used to localize sounds were binaural differences, then we might expect that we would often hear sounds that originated in front as coming from behind and vice versa. Since we are generally accurate in determining the locations of sounds, we can assume that there must be additional information available which enables the resolution of the ambiguities of binaural differences. This additional information is referred to as spectral cues.

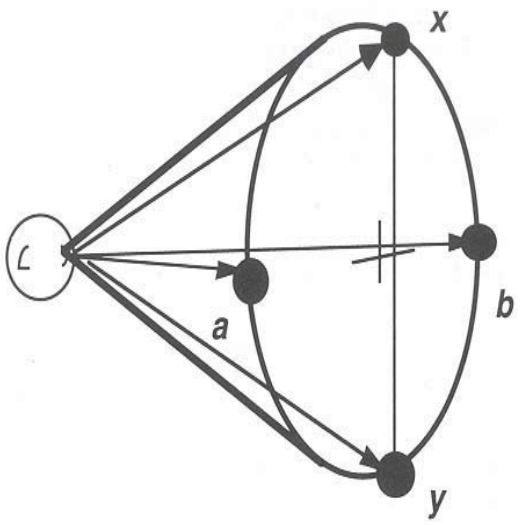


Figure 3: Cone of confusion. The axis drawn between the two ears is referred to as the interaural axis. A sound 30 degrees in front of the interaural axis (Point a) will have the same pattern of binaural differences as a sound 30 degrees behind the interaural axis (Point b). This is also true for a sound 30 degrees above the interaural axis (Point x) and 30 degrees below (Point y). If the head is assumed to be a sphere then the range of possible locations that are specified by any single binaural difference form a cone with the apex at the centre of the head (Reproduced from Begault, 1994).

As sound travels towards the ear it interacts with the torso, head and external ears (pinnae). This interaction causes a mix of direct and reflected sound reaching the ears. This mix results in some frequencies being amplified (causing spectral peaks) and others being attenuated (causing spectral notches). The end result is that we get a pattern of peaks and notches that are unique for a particular location in space (See Figure 4). We use these spectral cues to tell us whether a sound has occurred in front or behind, and where it lies in the vertical plane (i.e. its elevation). Thus spectral cues provide the extra information we need to help disambiguate binaural difference cues, and in combination these two sets of cues provide enough information to define a unique location in space.

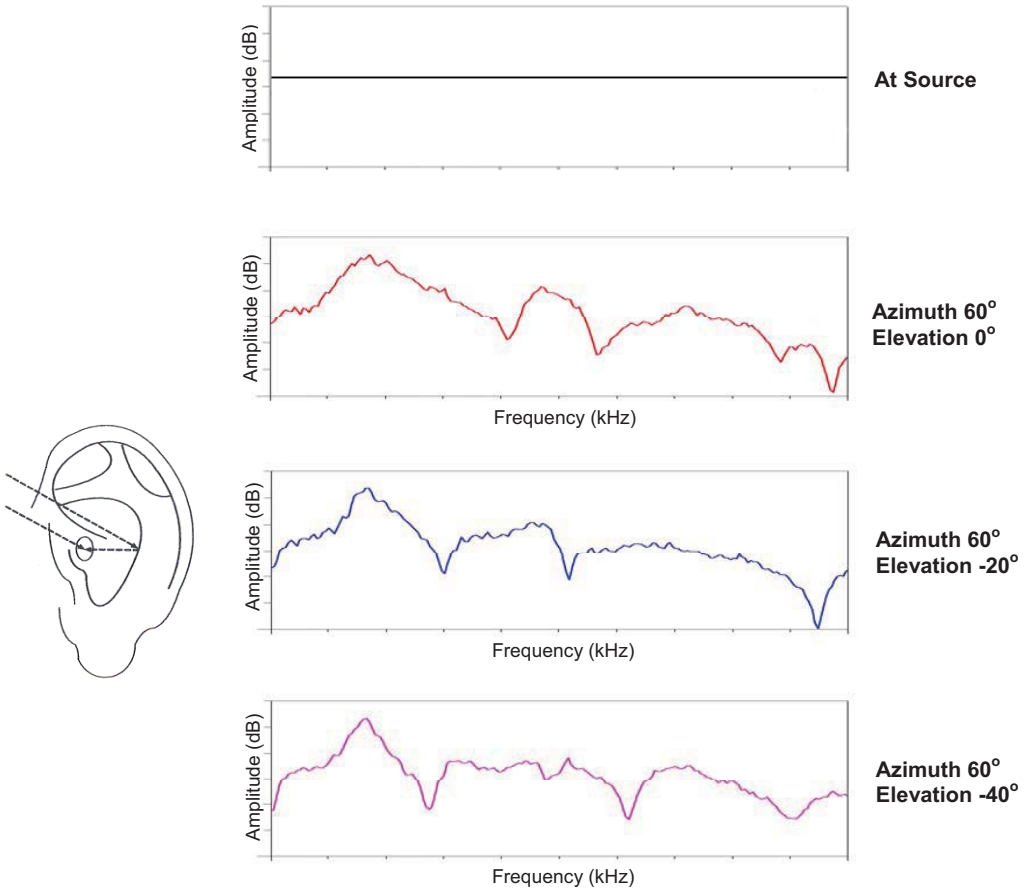


Figure 4: At ear canal recordings describing amplitude changes as a function of frequency at three different source locations (60˚ azimuth and 0˚, -20˚ and -40˚ elevation) demonstrating that each location has a unique pattern of peaks and notches. Also included is an at source recording describing the absence of any change when a sound does not interact with the external ears.

In addition to binaural differences and spectral cues there are other types of information we use to help localize sounds, one of these being head movement. Head movement provides an alternative method of disambiguating the location of a sound source. As we move our heads from side to side the pattern of binaural differences changes and the direction of this change tells us whether a sound is in front or behind. For example, if a sound is on the left side and we rotate our head to the right, then if the sound gets louder in the left ear it must be in front. Studies that have measured sound localization report that front/back and back/front confusion rates have been much lower when the head is free to move (Burger 1958).

All these cues together provide us with a library of information to use when making localization judgements. To produce synthetic 3-D audio the aim is to replicate all these

cues with a high degree of fidelity in order to create a situation where we mimic what would occur if a real sound had been played in the external world. If we faithfully reproduce all the normal cues we use, then someone listening to these cues should have exactly the same experience as if they were listening to a real sound.

3. Implementing 3-D audio

3.1 Techniques for implementing 3-D audio

The basis for producing 3D audio lies in the imitation of the cues necessary for spatial hearing. This means that the interaural intensity differences, the interaural time differences and all the relevant spectral changes need to be present in a signal before a sound will be heard at a required location. For example, for a sound to appear to originate 45 degrees in front of the interaural axis on the right side of the head you would require a signal delivered to the right ear louder than the signal delivered to the left ear, a signal delivered to the left ear delayed such that it arrived sometime after the signal reached the right ear, and spectral changes to both ears that accurately mimicked changes that would be expected if a real sound were presented at that location.

In virtual spatial hearing all of these changes can be imitated using Digital Signal Processing (DSP). Firstly, an accurate understanding is required of the signal changes that occur at each location to be represented in the virtual world. This provides the data for the second stage. In the second stage, once these changes are accurately identified, a pair of filters are constructed (one for each ear), that will accurately recreate all of these changes onto any signal passing through the system.

All the signal changes caused by binaural differences and the effects of sound interacting with the head torso and pinnae are referred to collectively as the Head Related Transfer Function (HRTF). It is the HRTFs that provide the data required to construct filters necessary to produce 3-D Audio.

3.2 Laboratory based 3-D audio

The first implementations of the techniques described in Section 3.1 were conducted in the laboratory. Some of the first researchers to create 3-D audio were Wightman and Kistler (1989a) who developed a technique based on the premise that "… if the acoustical waveforms at a listener's eardrums are the same under headphones as in free field, then the listener's experience should be the same" (Wightman and Kistler, 1989a, p. 859).

To record the HRTFs Wightman and Kistler used a probe microphone placed deep inside the listener's ear canal. The probe microphone consisted of a miniature microphone coupled to a silicone rubber tube. This tube was placed such that its tip was 1 to 2 mm from the listener's eardrum and was held in position using a customised shell that sat at the entrance of the listener's ear canal. Each recording consisted of the time domain representation of a signal recorded from each of the probe microphones (left and right ear)

in the listener's ear canal. This signal incorporated the directional filtering characteristics of the listener's pinnae, head and torso, but also the characteristics of the original test signal, the loudspeaker (or headphone) and the measuring microphone. To obtain an uncontaminated free-field-to-eardrum transfer function (HRTF) the authors removed the contaminants by dividing them from the HRTF in the frequency domain (Wightman and Kistler, 1989a). The recording procedure involved fitting the probe microphones to a listener and then positioning a speaker at a range of locations whilst recording the response to a broadband noise signal. Wightman and Kistler measured HRTFs with a speaker located at 144 different locations around the head of the subject, including elevations of –36º, -18º, 0º, 18º, 36ºand 54º and at azimuth locations right round the head separated by 15º.

In a subsequent study Wightman and Kistler (1989b) determined the accuracy of their 3-D audio by comparing real world (free-field) performance against virtual (3-D audio) performance. Perfect performance for the 3-D audio would be indicated by equivalence between an ability to tell where the sound was located in the real world and in the virtual world. Listeners were required to make absolute judgments as to the location of the sound source by calling out numerical estimates of apparent azimuth and elevation of the stimulus. These stimuli could come from any location between –36º to 54º in elevation and 360º in azimuth. Wightman and Kistler employed the same response methodology for both their virtual and free-field conditions. During the free-field condition the sounds were played from a speaker positioned at the required location. During the virtual condition the appropriate filters were used to recreate the apparent location of a sound at the desired location.

The major difference in performance reported by Wightman and Kistler (1989b) was an increase in the number of front/back reversals that occurred in the virtual (3-D audio) condition. A front/back confusion is said to have occurred when a stimulus located in front of the listener is perceived to have originated from behind or vice versa. The frequency of reversals was almost double for the virtual sources than for the free-field (Wightman and Kistler, 1989b). Another difference in localization performance between the virtual and real sound sources was in the vertical dimension, with vertical plane determination being slightly less well defined. Since both front/back and vertical plane discrimination are based on spectral information that arises from changes induced by the head, torso and pinnae, these results suggested that the spectral information was not being recreated with sufficient fidelity by the 3-D audio system.

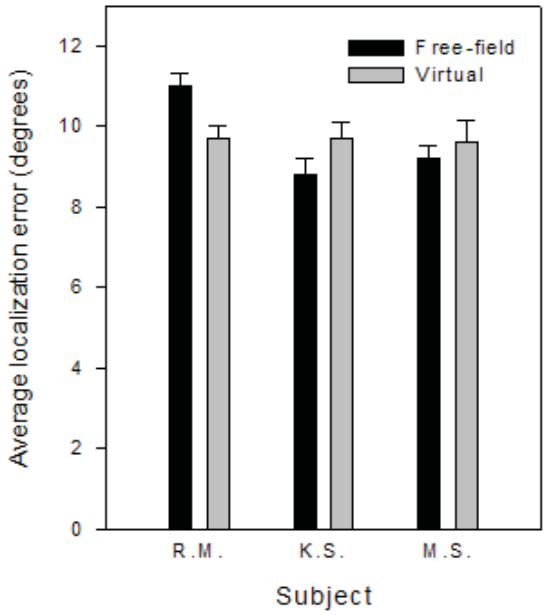
Other researchers have created 3-D audio using similar techniques with similar results. Pralong and Carlile (1994) used a variation of the technique described by Wightman and Kistler, (1989a). They used a probe microphone to make the recordings but implemented different techniques to create the customised shells used to keep the probe microphone in place. They chose to use the same moulding technique but plated the surface of the mould. This technique reduced the thickness of the mould achieved through electroplating so that it did not obstruct the path of the sound, which, the authors suggest, may have been a problem in earlier studies using this technique (Pralong and Carlile, 1994). In a separate study that evaluated the accuracy of the 3-D audio based on their recording technique,

Carlile et al (1996) reported a slight increase in front/back confusions and a slight increase in localization error in the virtual compared to the free-field condition.

Bronkhorst (1995) also used a probe microphone technique when measuring HRTFs. Listeners were seated in the centre of a sphere covered in speakers located in an anechoic chamber. Measurements were made from –56.3º to 90º elevation at 15º intervals in azimuth which resulted in 976 recording locations. Head position and orientation was measured using a magnetic head tracker. Feedback was provided to the subjects to ensure that their heads remained in the original reference position. If the subjects deviated from this position by 0.75 cm in the x, y and z directions, or 5º in azimuth or elevation, the measurement process was stopped until they regained their original reference position (Bronkhorst, 1995). To ensure that the frequency spectrum of the signal was flat when presented through the speakers, compensation was introduced into the stimulus. Bronkhorst reported increases in elevation error during the virtual listening conditions and sometimes large increases in front/back errors during virtual listening.

More recently techniques other than those using probe microphones have been used to make the initial recordings. Møller, Sorensen, Hammershoi, and Jensen (1995) used a blocked-ear-canal microphone placement technique. Miniature microphones were fitted into earplugs and then placed in the ear canal of the listener. The diaphragm of the microphone was positioned to be flush with the entrance of the ear canal (Møller et al, 1995). Measurement of HRTFs took place in an anechoic chamber inside which was a loudspeaker set-up consisting of eight speakers fixed in an arc of a circle ranging from –67.5º to 90º. Measurements were taken at every 22.5º around the head. In a subsequent study that continued the evaluation of the blocked-ear-canal microphone placement technique Møller, Sorensen, Jensen and Hammershoi (1996) concluded that the blocked ear canal method had a number of advantages including the ability to use larger microphones which allows better signal-to-noise ratios, a less invasive procedure that does not require insertion of probe tubes into the ear canal close to the eardrum, and a reduction in measurement variability associated with determining headphone transfer functions that allows for more accurate headphone equalization.

Martin, McAnally and Senova (2001) also used the blocked-ear-canal technique of microphone placement and fitted miniature microphones encased in swimmers' ear putty at least 1mm inside the listener's ear canal entrance. During the recordings listeners were seated in an anechoic chamber in the centre of a 1-m radius hoop on which a loudspeaker was mounted. The speaker was positioned at a range of locations between the elevations of –40º and 70º, separated by 10º intervals in the horizontal plane. The listener's head position and orientation was monitored using a magnetic head tracker with feedback provided to the listener as to their head position if it deviated from the original reference location. Martin et al (2001) report that for every participant in this study virtual localization was found to be equivalent to free-field localization performance across a large range of locations. This improvement in performance was attributed to procedures adopted to compensate for the filtering characteristics imposed on the signal by the microphones and headphones used, as well as techniques used to maintain the stability of the microphone during recording. The implication is that without proper compensation and recording techniques, imperfect 3-D audio will result.



Front/Back Confusion Errors

SUBJECT	FREE-FIELD	VIRTUAL
R.M.	4.8%	4.8%
K.S.	2.1%	3.0%
M.S.	1.8%	0.9%

Figure 5: Average free-field and virtual localization and front/back errors for each participant (reproduced from Martin, McAnally and Senova, 2001)

In summary, initial attempts to construct virtual audio reported higher front/back confusion rates during virtual listening conditions than those in the free-field, and an increase in errors for localization in the vertical plane. More recent attempts report equivalent performance between virtual and free-field listening conditions.

3.3 Individualized versus non-individualized HRTFs

One of the major sources of variation between different implementations of 3-D audio systems is the choice of HRTF set and as a result one of the major decisions required is whether the system will use individualized or non-individualized HRTF sets. This issue has been explored by researchers in the past, and continues to be explored, in an attempt to determine the relative impact of this choice under various conditions.

It is widely accepted that the use of non-individualized HRTFs will result in increases in virtual localization error. Wenzel, Arruda, Kistler and Wightman (1993) measured the virtual localization accuracy of 16 listeners using a non-individualized (sometimes referred to as generic) set of HRTFs, and compared performance against that measured in the free-field. These data were also compared to those from a previous experiment (Wightman and Kistler, 1989b) where listeners used individualized HRTF sets. Listeners were presented with stimuli from source locations all around the head. In all conditions participants were required to indicate the perceived direction of the sound by verbally responding with numerical estimates of azimuth and elevation (in degrees). In the free-field condition the sound was presented via a loudspeaker positioned at the appropriate

location, and subjects were blindfolded so that they could not see the speaker. In the virtual condition sounds were presented via headphones. Confusion rates were considerably higher for the virtual (31% of responses were front/back reversals) than for the real sources (19% front-back reversals). Elevation error also varied considerably, with many listeners exhibiting poor elevation accuracy in virtual conditions. The variability in elevation accuracy can potentially be explained by the degree of match between the listener's natural HRTF set and the non-individualized HRTF set used in the virtual condition. Listeners with a close match would exhibit a reasonable degree of elevation accuracy, whilst listeners that were not matched would exhibit much poorer elevation accuracy. Azimuth discrimination remained accurate under all conditions.

These results have been confirmed in a separate study by Begault and Wenzel (1993) who measured localization accuracy for listeners using non-individualized HRTFs with speech signals rather than the broadband noise signal used in Wenzel et al (1993). They again reported confusion rates of 29% (similar to the 31% reported in Wenzel et al, 1993), increased elevation error (mean 17 degrees) and accurate azimuth discrimination.

Bronkhorst (1995) noted that the increase in confusion rates was "a disturbing factor in practical applications, especially when short sounds are presented and fast reaction times are required" (Bronkhorst, 1995, p. 2543). He investigated differences between individualized and non-individualized HRTFs used to create virtual sources in an attempt to determine the degree to which localization performance depends on listener-specific cues. Bronkhorst concluded that virtual sound sources created using individualized HRTFs could be localized accurately provided that head movements were allowed and that the sound was long enough to allow the use of head movements. Localization performance using non-individualized HRTFs was much more variable and generally poorer. The extent of the degraded performance was difficult to determine due to the high confusion rates exhibited under all conditions including real sound source conditions.

A further study by Møller et al (1996) reached similar conclusions. Localization performance was studied when subjects listened to real sounds, virtual sounds using individualized HRTFs and virtual sounds using non-individualized HRTFs. Eight subjects listened to sounds presented from 19 locations distributed around the subjects such that they varied in azimuth, elevation and distance. Subjects could see the speakers and responded on a digitising tablet by indicating the speaker location where the sound appeared to originate. Møller et al (1996) reported that there was no difference in localization accuracy between conditions when subjects were listening to real sounds and when subjects were listening to virtual sounds using individualized HRTFs. They did report significant differences when listening to virtual sounds using non-individualized HRTFs with increased errors particularly in the median plane where there were increased numbers of front/back confusions.

There is therefore a body of research that suggests that the use of non-individualized HRTFs results in an increased number of front/back confusions and an increase in elevation error, whereas azimuth error remains accurate. While this is generally recognised, it is often argued that these results have been collected in the absence of head movements, and that when head movements are allowed it provides additional

information that resolves front/back ambiguities, reducing the difference between localization accuracy using non-individualized and individualized HRTFs.

Some preliminary work at DSTO (McAnally and Martin, 2004) suggests that whilst head movements can resolve front/back confusions, it would appear to be at some cost. While the number of front/back reversals was reduced to zero, other measures suggested that response time was longer and that under some conditions a substantial head movement was required before the front/back confusion was resolved. Consideration of the mechanism for resolving front/back confusions gives a hint as to why an increase in response time is to be expected. When a sound is heard, if the listener has to wait to initiate head movements before being able to identify the correct location of the sound, then there will be an automatic delay whilst head movement is initiated and the resultant information interpreted. This delay can be increased if the subject is attempting to visually locate a target and if the listener rotates his/her head in the wrong direction moving away from the target rather than closer. McAnally et al indicated that there was only marginal improvement in elevation accuracy even with head movements.

Whilst these results are preliminary and further experimentation is required, it does suggest that the requirements of the task to be augmented by 3-D audio may need to be considered with care. For tasks where response time is critical and/or location accuracy important, there may be a case for requiring the best fidelity HRTF available in order to minimise the number of front/back confusions. At present this would require the use of individualized HRTFs until improvements can be implemented for non-individualized HRTFs. For tasks where response time is not critical and/or location accuracy not as important (e.g. communications segregation), designers may have a broader choice in the selection of HRTF type.

It is important to note that the debate regarding whether individualized or non-individualized HRTFs are the most appropriate choice has no real impact on the design of a 3-D audio system. In later sections of this report the requirements for a 3-D audio system are listed, and one of the primary requirements identified is for a capability to import and store an HRTF set. The exact nature of the HRTF set and how it was generated is irrelevant to the system requirements. The requirement remains the same regardless of whether the HRTF set is individualized, non-individualized, recorded using an artificial head or generated using mathematical algorithms. The end result is still an HRTF set, and as long as a 3-D audio system is designed appropriately the method used for generating the HRTF set can change and the end product can still be imported into the system.

3.4 Alternative techniques for generating HRTF measurements

Whilst making HRTF measurements for each listener is an accepted practice in the laboratory where accuracy is a primary concern, there are a number of reasons why this is not a convenient procedure for commercial applications. This is particularly true where the application is aimed at a mass market where making individualized recordings for every user is neither feasible nor desirable. For this reason there is ongoing research investigating alternative methods of generating HRTF measurements which allow the listener to experience 3-D audio without the need to make HRTF measurements for each

individual listener. Such methods include the construction of geometric models for generating HRTFs, using artificial heads to make recordings, predicting HRTF sets based on the morphology of the individual and making adjustments to generic HRTF sets so that they are a closer match to the individual.

3.4.1 Geometric models

Most 3-D audio systems using non-individualized HRTFs use a single HRTF set (recorded on one individual) for all users. The accuracy associated with this approach is therefore dependant on the match between the physical characteristics of the user and those of the individual associated with the HRTF recordings used. One method for generating an HRTF set that is more generic in nature is by mathematically modelling the signal changes caused by the head, torso and pinnae in an attempt to extract those fundamental features common to all potential users. Such an approach has been adopted in the construction of the HEAD Acoustics system (see Figure 6) where those parts of the head identified as the most acoustically relevant are modelled by a sphere and two elliptic disks with an eccentric cylindric cave. HRTFs can be derived directly from these geometric parameters that include the diameter of a sphere, dimensions of disks and cave, direction of source, etc. In order to ensure that the modelling calculations are kept to a manageable amount, it is necessary to minimise the number of geometrical parameters used. The end result therefore is a simplification of an HRTF set which ideally contains the most critical features required for accurate localization. This approach has resulted in the generation of a reasonable 3-D audio experience, although the reductionist approach results in a corresponding loss of acuity in the vertical plane and some increase in the number of front/back errors. It might be expected however that with further refinements to the geometric models used, improvements in the 3-D audio experience may well be observed.

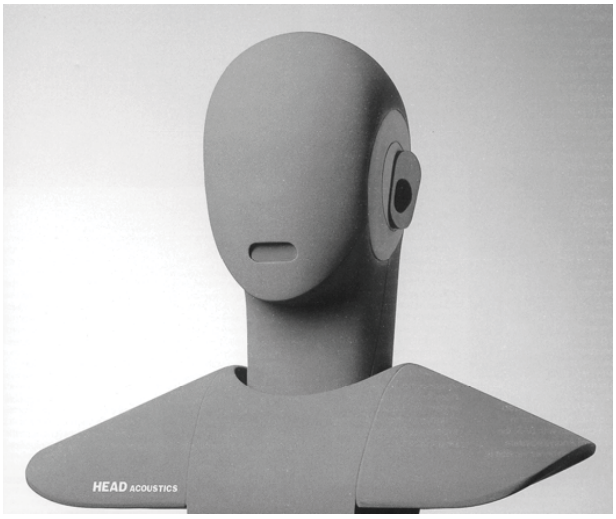


Figure 6: Artificial head manufactured by HEAD Acoustics which includes geometric pinnae designed to model the principal physical features of human listeners

3.4.2 Artificial heads

As mentioned in Section 3.1 the creation of 3-D audio requires as a first stage the measurement of changes that occur to an audio signal as it travels to the eardrum of the listener. There are numerous alternative methods for generating these measurements. The first attempts to construct 3-D audio made recordings using the actual listeners. Such a process requires fitting microphones into the ears of the listener, having them seated in the centre of a speaker arrangement for generating sounds whilst recordings are made of the responses from the microphones. It is not always convenient to make recordings in the ears of actual people, particularly when large numbers of listeners are involved or large numbers of locations are required (e.g. where recordings are required every 1^{o} all around the head).

As an alternative to generating measurements in real individuals, researchers have examined the possibility of using artificial heads during the recording process. The advantage of using artificial heads is that recordings are repeatable and large numbers of locations can be recorded easily. The difficulty with such an approach is that the recordings are non-individualized and there is some loss of localization acuity. Minnaar, Olesen, Christensen and Møller (2001) assessed localization accuracy for binaural recordings from 10 different artificial heads and two human heads. The artificial heads included commercially available systems from Knowles Electronics (KEMAR), Neumann (KU100), Bruel and Kjaer (4100), Head Acoustics, Cortex Electronic (MK1) as well as artificial heads developed in university research laboratories such as the Aachen University ITA and the Aalborg University VALDEMAR.

Binaural recordings were made using each of the artificial heads and in the two human heads and played back to 20 listeners with sounds recorded from 19 different loudspeaker locations distributed around the listener. The recordings were presented to the listeners over headphones randomly, one speaker location at a time, after which the listener was asked to indicate which of the 19 potential locations the sound appeared to originate from. The results indicated that there was an increase in the number of errors associated with recordings made in the artificial heads, applying to all heads and nearly all directions. The lowest number of errors occurred for recordings made on the two human heads. Among the artificial heads, VALDEMAR and ITA gave the lowest number of errors. The results suggested that localization performance with binaural recordings is primarily influenced by properties of the recording heads and that the artificial heads differed significantly from one another. The implications of this research are that any intention to use artificial heads in the recording process should consider the choice of recording instrument carefully. The relative performance associated with the artificial head VALDAMAR can be attributed in part to the design features of the manikin. The head and the torso were designed based on acoustical measurements on 40 humans and from anatomical data. The pinnae were casts of a human pinna from a subject used in earlier studies who had demonstrated localization proficiency. Thus whilst VALDEMAR was an artificial recording instrument, its design closely resembled an actual human with some attempt to represent average dimensions.

In summary, whilst using an artificial head is probably the simplest method for conducting the recordings necessary to implement spatial audio, a major problem is that the recordings generated using this method are not individualized. More recent research has begun examining methods that can be used to generate an HRTF more representative of the individual listener. Two such approaches include either firstly adjusting existing recorded HRTFs in such a way as to make them more similar to those used by the listener, or secondly instead of recording the HRTFs, using predictive algorithms based on morphology to artificially generate a set of HRTFs that more closely resemble those of the listener.

3.4.3 Adjusting generic HRTF sets

Previous studies have demonstrated that the major source of inter-listener differences in HRTF sets is not in the low-frequency interaural time and intensity difference cues (ITDs and IIDs), but in the high-frequency spectral cues (Jin, van Schaik, Best and Carlile, 2003). The implication here is that approaches aiming to adjust generic HRTF sets should adapt the high-frequency spectral cues of the sound for each listener in order to achieve improved fidelity. Middlebrooks, Makous and Green (1989) mapped the relationship between the physical size of the listener (overall height) and the frequencies at which listeners exhibited directional sensitivity. The nature of the relationship was such that the frequencies of maximum directional sensitivity were in the higher frequencies for smaller subjects and in the lower frequencies for larger subjects. This means that lower-frequency sounds with longer wavelengths tend to interact more with larger heads and ears, while higher-frequency sounds with shorter wavelengths tend to interact more with smaller heads and ears. These observations lead Middlebrooks et al to conclude that it might be possible to reduce inter-listener differences by scaling the HRTFs in frequency, which might in turn lead to improvements in 3-D audio accuracy when using non-individualized HRTFs. They examined whether HRTFs differed systematically among listeners in regard to the frequencies of spectral features, as well as whether an optimal scale factor existed between the HRTFs of different listeners that could be predicted by the relative physical sizes of the individuals.

In a later study Middlebrooks (1999a) again examined individual differences in external-ear transfer functions in order to identify systematic inter-listener differences in HRTFs that are critical for accurate localization. He measured the directional transfer functions of 45 subjects to compare and contrast spectral features. Each HRTF was treated as a combination of two transfer functions. The first transfer function contained those components that varied as a function of sound source direction. The second transfer function contained components that were common to all sound-source directions. Contributors to the common components (non-directional) included the canal resonances and the microphone and headphone transfer functions. To isolate the components that changed as sound source direction altered, the common components were removed from each HRTF so that only the directional components remained. These altered HRTFs are referred to as the *directional transfer functions* (DTFs; Middlebrooks and Green, 1990). The DTFs of the 45 listeners showed patterns of spectral features (e.g., spectral peaks, notches, and slopes) that varied systematically as a function of sound-source elevation and azimuth. For example, as sound sources were shifted in elevation from low to high, the

centre frequencies of spectral features tended to shift from low to high frequencies. Whilst the DTFs contained similar patterns of spectral features, the relative position of those features in terms of frequency tended to vary from listener to listener. For example, the spectral features for listener S35 tended to occur at higher frequencies than those for listener S07 (See Figure 7). Middlebrooks was able to demonstrate that by applying a scaling factor and shifting the frequency locations of spectral features, it was possible to align those spectral features between two listeners and therefore reduce the inter-subject variability. The implications of such an approach are that if a non-individualized set of HRTFs were used in a 3-D audio system, and the spectral features of that non-individualized set could be shifted in the frequency domain so that they aligned more closely to those of the listener, it may result in more accurate 3-D audio.

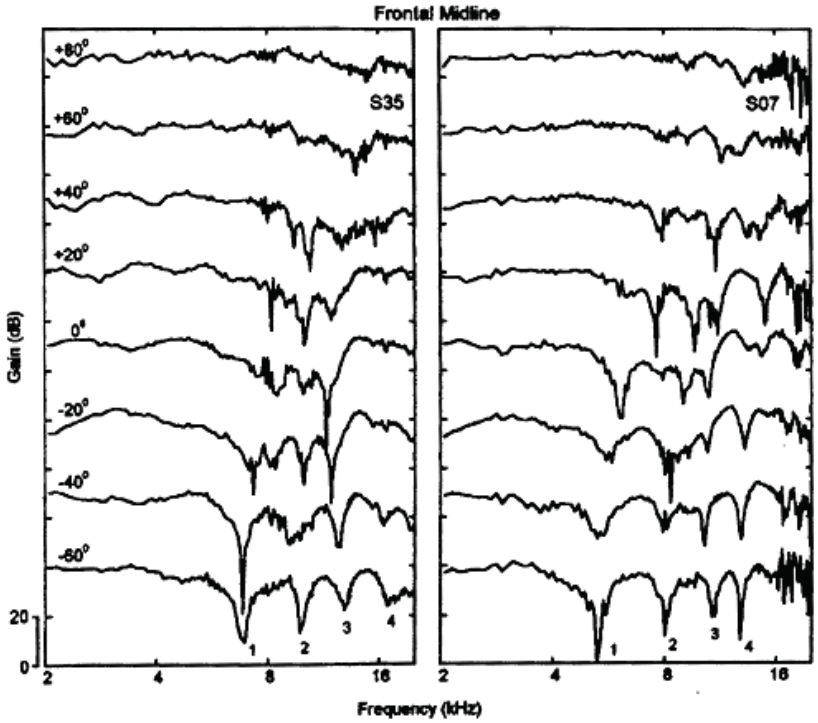


Figure 7: Directional transfer functions (DTFs) for directions on the frontal midline at the elevations indicated relative to the horizontal plane. Left and right panels show DTFs measured in the right ears of listeners S35 and S07 respectively (reproduced from Middlebrooks, 1999a).

In a follow-on study Middlebrooks (1999b) was able to demonstrate that by applying frequency scaling to non-individualized DTFs and reducing spectral differences between the listener and the non-individualized HRTF set, improvements in virtual localization were evident. The implication is that it may be possible to improve the performance associated with using an HRTF set that is not your own. In this study Middlebrooks examined virtual sound localization in three conditions that differed according to the directional transfer functions (DTFs) that were used to synthesize virtual sounds. In the

first condition listeners used DTFs as measured in their own ears. In the second condition listeners used DTFs measured in someone else's ears. In a third condition listeners used DTFs measured in someone else's ears but shifted in frequency to minimise the mismatch between the spectral features of the listener and the spectral features of the DTFs from another listener. Virtual localization performance was best when listeners were using DTFs from their own ears. In other-ear conditions errors tended to increase in proportion to the inter-subject differences in DTFs. When spectral features in the other-ear set of DTFs fell systematically lower in frequency than in a listener's own DTFs, targets would typically be reported with a downward bias (targets were perceived as lower in elevation than the actual location). When spectral features in the other-ear set of DTFs fell systematically higher in frequency than in a listener's own DTFs, elevation judgments showed an upward bias (targets were perceived as higher than the actual location). When scaling was applied such that the spectral features associated with the DTF being used were moved such that there was closer alignment with the spectral features of the listener, all measures of performance tended to show improvement.

The work of Middlebrooks indicates that it may well be possible to customise generic sets of DTFs to individual listeners. He suggests that there are two ways in which one might improve virtual localization with non-individualized DTFs. The first possibility would be to maintain a single set of DTFs from one listener and to adapt those DTFs to individual listeners by scaling them in frequency to minimise any differences. The second possibility would be to maintain sets of DTFs from multiple subjects and to select a DTF that was similar to the listeners own DTF. The difficulty with either of these possibilities is that it requires detailed knowledge of the listeners DTFs, which generally requires making acoustical measurements on the listener. As Middlebrooks indicated, it may be possible to make estimates based on physical measurements of the width of the listener's head and height of his/her external ear to avoid the necessity of making recordings.

These results should be interpreted with care however, for whilst the techniques used by Middlebrooks improved virtual localization performance when using generic DTFs, there was still a considerable difference in performance compared to when listeners used their own DTFs. This research therefore indicates that there is a distinct possibility that in the future, acoustic measurements for individual listeners may not be required and that techniques may be available to adapt generic HRTFs to the listener. Until such time that these new techniques can demonstrate equivalent virtual localization performance, tasks that require high fidelity 3-D audio may well need to continue to use individualized HRTFs. Fortunately, as indicated earlier, it is possible to design a 3-D audio system that will accommodate both sets of requirements.

3.4.4 Morphology

The difficulty with either of Middlebrooks' possibilities for customising generic sets of DTFs to individual listeners is that it requires detailed knowledge regarding the DTFs of the listeners. This negates one of the primary advantages of using generic HRTF sets which is the lack of the requirement to make acoustical measurements on each individual listener. Middlebrooks did however indicate that there may be methods for making

estimates based on physical measurements of the width of the listener's head and height of his/her external ear (i.e. the morphology of the listener).

The spectral details of a DTF result from the physical interaction of the incoming soundwave with the head, torso and external ears. It is not surprising therefore that the frequencies of spectral features are related to the size of the head and the external ears. Middlebrooks (1999a) measured a set of physical dimensions on 33 individuals and generated a set of DTFs for each individual. A number of points on the external ear were measured including the pinna-cavity height as measured from the inter-tragal notch to the rim of the helix (See points I and A respectively on figure 8). In addition some measurements of the head were included, as well as overall body height. Inter-subject ratios of physical dimensions were compared with inter-subject optimal scale factors, with indications that the inter-subject optimal scale factors could be estimated with some accuracy from ratios of certain physical dimensions. The pinnae-cavity height and the head width both showed strong correlations with optimal scale factors.

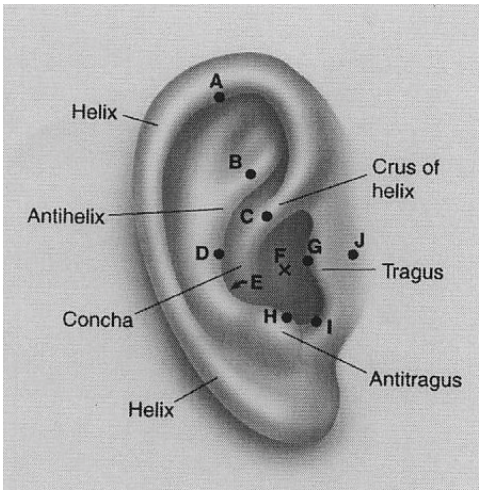


Figure 8: Diagram of the features of the external ear. Dots labelled with letters indicate points at which physical measurements were taken by Middlebrooks, 1999a. Point A is located on the uppermost point of the rim of the helix and point I is on the ridge of the crus that maximised the distance to the bottom of the inter-tragal notch (reproduced from Middlebrooks, 1999a).

Other researchers have also begun to address this issue. Bronkhorst (1999) argued that adaptation of HRTFs to an individual could be performed by reconstructing a set of HRTFs using a limited set of parameter values. He listed two methods for estimating these values as firstly, prediction based on anthropometrical data, and secondly, optimisation using a simple listing test.

A similar approach has also been adopted by Jin, Leong, Leung, Corderoy and Carlile (2000) who used a generative statistical model of DTFs for a population of 36 people to provide a basis for systematically varying the degree of matching between test DTFs and

true, individualized DTFs. This allowed them to examine the sensitivity of human sound localization performance to individual differences in DTFs, as well as to examine the mapping between the morphology of the external ear and individualized DTFs. The results indicated that listeners were sensitive to approximately 60% of the individual variations in the population HRTFs, and that sound localization performance maintains some degree of accuracy even when accounting for only 30% of individual variations. The comparison between morphological features in the DTFs was performed showing some degree of correlation between spectral features in the DTFs and morphological landmarks. Jin et al conclude that functional mapping between morphology and DTFs appears to be feasible.

3.4.5 Summary

Whilst some progress has been made in identifying alternative techniques for generating HRTF measurements, none of the techniques so far have resulted in accuracy similar to that observed when individualised HRTFs are used. Depending on the intended application for the 3-D audio system, for situations where absolute accuracy is required the optimal choice still remains the use of individualised HRTFs. In terms of the design of a 3-D audio system, one of the fundamental requirements for the implementation of 3-D audio is that there be a capability to load in HRTF filter sets. The method of generating the filter sets is independent of this requirement and can be periodically reviewed as better methods become available. As long as there is some degree of flexibility in the system with regards to the structure and composition of the HRTFs, improvements in techniques for generating HRTFs will be able to be accommodated as they occur.

3.5 Remaining issues

This section includes a number of issues that need to be considered when implementing 3-D audio. Many of these are the subject of ongoing research, and manufacturers are addressing some of these issues. Whilst most of these issues do not necessarily have a major impact on the design of a 3-D audio system, each of these factors needs to be considered in order to ensure successful implementation.

3.5.1 Sound content

Not all sounds are appropriate for use in 3-D audio systems. Inappropriate sounds will result in large errors (e.g. loss of front/back discrimination and/or elevation discrimination) with a corresponding poor 3-D audio experience. In such cases it does not matter how well designed the 3-D audio system is, as the system will be incapable of overcoming the deficiencies of the sounds used and will perform poorly because of inappropriate sound content.

The majority of sounds that have been used in laboratory evaluations of 3-D audio have comprised of some form of noise signal. As described in Section 3.4.3 binaural time difference cues are mostly associated with low frequencies and spectral and binaural intensity difference cues are mostly associated with high frequencies, so a broadband

signal provides for both sets of cues. Noise signals however are not readily identified, discriminated or contain much meaning, and are therefore not appropriate for use as auditory warnings. What is required is a set of clear guidelines that allow the design and construction of sounds that will function effectively as auditory warnings as well as contain sufficient spatial information to function as 3-D audio sounds.

Some work has been conducted in this area. We know for example that auditory signals must be broadband to some extent, and must contain energy above about 4 kHz in order to be localized accurately (e.g. Butler and Planert, 1976; Hebrank and Wright, 1974, King and Oldfield, 1996). This is supported by studies that have examined how well speech sounds can be localized, where it has been observed that speech segments containing high frequency content (e.g. fricatives) are localized better than speech segments containing only low frequency content (e.g. vowels) (e.g. Davis and Stevens, 1974; Shigeno and Oyama, 1983).

Some work has been conducted to identify techniques that can be used to add high frequency content to existing sounds to improve their performance when used in 3-D audio systems. Several years ago the Defence Research Agency (DERA) developed a set of 12 auditory warnings for use in military helicopters as warnings for various flight systems. The original signals consisted of frequencies predominately below 4 kHz that made them unsuitable for use in 3-D audio systems. Patterson and Datta (1994) examined methods for increasing the frequency range of the DERA warnings up to 12 kHz whilst minimizing any other changes that might alter the recognisability of the warnings since they were already installed in operational aircraft. Patterson and Datta observed that since the temporal pattern of a complex sound was a major component of the character of a sound, and that since the main result of adding high frequency content to a sound was to brighten the timbre, it should be possible to add high frequency content using the same temporal envelope. Patterson and Ditta identified a number of techniques that would fulfil these requirements including envelope filling, nyquist whistling and fine structure doubling.

Martin, Parker, McAnally and Oldfield (1996) evaluated the localization accuracy of a representative subset of five of the original 12 auditory warnings from DERA and the corresponding high frequency versions produced by Patterson and Ditta in order to determine the relative success of the various techniques used to add high frequency content. Of the three modification methods employed, only one (fine structure doubling) produced changes to the warnings that improved localization accuracy when compared to the original versions of the warnings. Whether this technique is applicable to other types of complex sounds remains to be determined.

Other work suggests that not every type of broadband signal is useful, and that it may be more important to have a spectrum that contains some constant bands with high amplitude (Blauert, 1969/70; Sextant, 1997). In a comparison of different examples of complex sounds, Griffiths, Watt and Parker (2005) found that sounds that had substantial high frequency content were the most accurately localized, but that those with more dynamic change in the high frequencies (e.g. repeating sections with sharp rise/fall times) performed better than those with continuous components. These studies suggest that

some degree of complexity appears to be preferable, although this observation needs to be verified by further research.

Despite some attempts to determine the requirements for sounds to be used in 3-D audio systems, there is in general a paucity of research on the topic and much more work is required to ensure that sounds for use in 3-D audio systems can be designed with confidence.

3.5.2 Background noise

Aviation platforms in general have high levels of background noise and one issue that needs to be considered is the impact that this background noise might have on the successful operation of a 3-D audio system. Most existing work that has examined this issue has evaluated the noise signal against a noise background (e.g. Good and Gilkey, 1996). The difficulty with this approach is that whilst the results indicate some loss in localization ability in the presence of background noise, it is unclear how much is due to some form of interference effect caused by the presence of noise and how much is due simply to an inability to detect the presence of a noise signal against a noise background. Further work needs to be conducted to establish the relative impact when more appropriate audio signals are evaluated which would make the detection of the signal more obvious against an appropriate noise background.

There are various approaches that can be used to reduce the level of background noise including passive attenuation (i.e. earplugs), Communication Ear-Plugs (CEP) and Active Noise Reduction (ANR). The implication for the design of a 3-D audio system is that the design may need to reflect the presence of one of these approaches. For example, if CEP is being used the fidelity of the transducers in the plugs may need to be considered to ensure that the CEP system is capable of producing the fidelity required (e.g. stereo, high frequency reproduction). The most suitable approach would be one that integrates noise reduction techniques with the 3-D audio system. An example of such an approach is the Terma system that is examining the use of ANR in conjunction with a 3-D audio system.

3.5.3 Hearing loss

As a function of operating in environments with high levels of background noise it is not uncommon for aviation operators to suffer some degree of hearing loss. The exact nature of this hearing loss and the severity of such loss is not well documented. Hearing tests in the aviation industry tend to focus on the lower frequencies (≤ 8 kHz) and are designed to identify substantial losses that have implications for the perception of speech. There are a number of studies that have recorded some loss in localization performance in the presence of hearing loss (e.g. Noble, Byrne and Lepage, 1994). The relative impact of milder hearing loss is less well understood and needs further research. It may cause some loss in localization performance, but how much and what impact is as yet not known. One possibility however is that the use of individualized HRTFs may allow minor adjustments to be made within the HRTF to compensate for very minor hearing deficiencies. Again, this is an issue that would need verification.

The immediate impact of this issue for the design of a 3-D audio system is negligible, however as for other factors, what is required is some form of systematic evaluation. The nature of any hearing loss needs to be assessed, the relative impact on the performance of 3-D audio evaluated, and the potential to compensate for any minor loss determined.

3.5.4 Appropriate application

Another issue to be considered in the implementation of 3-D audio is the number of potential applications for its use. Whilst 3-D audio is often mentioned in the EW context, other potential uses for 3-D audio abound. The main systems that have been explored as possible sites for the implementation of 3-D audio can be divided into three; Electronic Warfare (EW) systems, Internal Communication Systems (ICS), Helmet Mounted Display (HMD) systems. This is not an exclusive list of potential systems and others are possible, but this list represents those that have been considered thus far and where some degree of technical development has been conducted.

Because of the obvious advantages of providing location information for critical threats EW systems have been one of the first to be considered. There are well-developed solutions for this option (e.g. Terma) and opportunities in the Australian context (e.g. BAE Systems) that make this an attractive option. The disadvantage of confining a 3-D Audio capability to an EW system is the potential inability to broaden the use of 3-D Audio to systems other than EW (e.g. communications to segregate radios). The implementation of 3-D Audio within an EW system does not preclude the use of 3-D Audio in other systems, but the application of 3-D Audio to a broad range of uses would be more readily achieved if it were implemented in other systems (i.e. the ICS or HMD systems).

4. Applications of 3-D audio

Apart from research examining issues associated with the design and construction of 3-D audio (i.e. how best to build it and make it work), there have been a small number of studies that have examined potential applications for 3-D audio (i.e. how best to use it).

4.1 Target acquisition

This body of research had its origins in laboratory studies that demonstrated clear benefits of providing spatial location information during a range of tasks. Perrott, Cisneros, McKinley and D'Angelo (1995) reported that search times for visual targets were 10-50% faster when those targets were paired with a free-field sound coming from the same location. This advantage was most obvious for targets behind the head, but response times were also faster for auditory-cued targets located in front. Similar benefits have been demonstrated under virtual listening conditions. Flanagan, McAnally, Martin, Meehan and Oldfield (1998) reported improvements in performance of a visual search task when 3-D audio is provided in addition to a Helmet-Mounted Display in a virtual environment. Bolia, D'Angelo and McKinley (1999) reported that providing either free-field or 3-D audio

cues resulted in decreased search times for a visual search task, and that the advantage of providing auditory cues increased as the number of visual distracters increased.

The effectiveness of supplementing Head-Down Displays (HDDs) with high-fidelity 3-D audio was investigated by Parker, Smith, Stephan, Martin and McAnally (2004) using a flight simulation task in which participants were required to visually acquire a target aircraft. There were three conditions: a visual HDD providing azimuth information combined with a non-spatial audio cue, a visual HDD providing azimuth and elevation information combined with a non-spatial audio cue, and a visual HDD providing azimuth information combined with a 3-D audio cue. Participants were asked to follow a lead aircraft and at some stage during the flight a warning was generated indicating the location of a target aircraft. Participants were asked to visually acquire the target aircraft and to position a sighting reticule over the image of the target aircraft. The position of the sighting reticule was slaved to the orientation of the participant's head via a magnetic head-tracker. When 3-D audio was presented, visual acquisition time was faster (by almost 4 secs), perceived workload was reduced and perceived situational awareness was improved. This performance improvement was attributed to the fact that participants were often able to perform the task without the need to refer to the HDD, enabling more time to be spent looking outside the cockpit searching for the target.

4.2 Communications segregation

Communications are an important part of any military aviation environment. It is not uncommon for aircrew to be required to monitor multiple radios simultaneously. For example, aircrew can be required to monitor internal communications between pilots and loadmasters, communications from tactical support units, other aircraft in formation, communications from headquarters and communications from tactical air controllers. In existing communications systems all these incoming communications are presented diotically (i.e. where all radio traffic is routed through a single channel and presented to both ears simultaneously). During diotic listening conditions when more than one message is received simultaneously it can be confusing to the listener, which in turn may lead to reduced intelligibility. Unfortunately in many cases, critical situations are paired with increased radio traffic, and it is possible that aircrew can miss important messages at times when they can least afford to.

For some time now there has been mounting evidence that spatial separation of simultaneous voices can improve speech perception. This observation was first noted by Cherry (1953) who observed that in our everyday natural environment it is relatively easy to understand one talker even if other people are talking at the same time. This ability has since been termed the "cocktail party effect", and a large amount of research has been devoted to understanding how this is achieved (for a review of research on the cocktail party effect see Yost, 1997, and Bronkhorst, 2000). Research investigating the cocktail party effect has identified a number of factors that contribute to this ability, one of which is the spatial separation of simultaneous signals. Other factors also have some influence (e.g. pitch, attention, etc.) but our interest in this report is on the factor of spatial separation.

Following on from the work by Cherry a number of studies have demonstrated substantial performance improvements when using virtual audio displays to spatially segregate the apparent locations of competing communication signals. These studies have used a variety of different techniques to generate the spatial separation ranging from generic binaural listening techniques to more involved 3-D audio processing which included HRTFs. Early studies examined the impact on intelligibility when the target speech signal was presented in the presence of some form of general background masking sound.

Bronkhorst and Plomp (1992) made recordings of short sentences in frontal positions against a noise background using an artificial-head (KEMAR) to simulate situations with competing talkers. They evaluated intelligibility by measuring the speech-to-noise ratio required for 50% intelligibility (referred to as the Speech-Reception Threshold (SRT)) and reported improvements in intelligibility as the noise maskers moved from all together in the front, to separate positions around the listener (i.e. as separation between the target signal and the competing background masker was introduced). Other studies which have used synthetic 3-D audio to achieve spatial separation report similar results. For example, Begault and Erbe (1994) who used non-individualized HRTF filtering for spatializing four-letter words (i.e. "call signs") against a background of multi-talker bubble. They reported significant improvements in intelligibility as the target was separated from the background speech babble and was presented at locations to the side of the head. These results were repeated in a subsequent study where the background was speech noise (Begault, 1995). Ricard and Meirs (1994) examined the intelligibility of speech and the ability of listeners to locate speech when direction information was added to the signal. They measured the impact of separation in the horizontal plane in the presence of masking noise using HRTFs to provide spatial separation. Ricard and Meirs found an improvement in intelligibility as the target signal was separated from the background masking noise.

A number of studies have examined how intelligibility is affected when there is a requirement that each talker be intelligible in the presence of multiple talkers presented at different (virtual) positions. Crispien and Eherenberg (1995) measured the ability of subjects to discriminate multiple simultaneous speech stimuli (short sentences) when presented from four separate locations and found the intelligibility scores for words was on average 51%. In a simulated cocktail party situation, Yost, Dye and Sheft (1997) used speech (words) uttered by up to three simultaneous talkers, and presented from seven possible loudspeakers in a semicircle in front of either a human listener, a single microphone, or an artificial head (KEMAR). With three concurrent talkers, average word intelligibility was around 40% for live listening and when listening to KEMAR recordings, compared to only 18% for the single microphone condition without any location information. Ericson and McKinley (1997) measured sentence intelligibility for two and four concurrent talkers against a background of noise where listeners had to reproduce sentences that contained a certain call sign. Three conditions were tested including diotic (where the sound delivered to each ear is the same e.g. two messages both sent to each ear), dichotic (where the sound delivered to each ear is different e.g. one message sent to the left ear and a different message sent to the right ear) and directional presentations (KEMAR recordings in the horizontal plane). With two talkers, scores were more than 90% correct for both dichotic and directional presentation when the two talkers was separated by at least 90° in azimuth. With four talkers and low-level noise, the advantage for

directional over diotically presentation was around 30% with 90° separation between talkers.

Drullman and Bronkhorst (2000) examined the possible merits of using three-dimensional auditory displays to improve not only intelligibility but talker recognition against a background of competing voices. They used words as well as sentences, and varied the number of competing talkers from one to four, whilst altering the virtual position of the talkers in 45° steps around the front horizontal plane. Compared to monaural and binaural presentation, 3-D audio presentation yielded better speech intelligibility with two or more competing talkers, particularly for sentence intelligibility. Talker recognition scores were also higher for 3-D audio than for monaural and binaural presentation although the differences were small. For binaural and 3-D audio presentations, the time required to correctly recognise a talker increased with the number of competing talkers. With two or more competing talkers, 3-D audio presentations required significantly less time than the binaural presentations. It is also worth noting that Drullman and Bronkhorst compared the use of individualized versus non-individualized HRTFs and found no major difference. The authors mentioned however that absolute localization was relatively poor which makes a comparison between individualized and non-individualized HRTFs difficult.

More recently, the spatialisation of speech signals has been shown to help operators function in complex auditory environments. There have been studies that have demonstrated an enhancement of speech intelligibility in the horizontal plane (Nelson, Bolia, Ericson and McKinley, 1999), vertical plane (McAnally, Bolia, Martin, Eberle and Brungart, 2002), and when talkers are spatially separated in distance (Brungart and Simpson, 2002). Studies have also demonstrated that spatialization of speech signals can compensate for the degrading effects of noise (Ericson and McKinley, 1997; Ricard and Meiers, 1994) and lower the perceived mental effort associated with attending to simultaneous speech streams (Bolia, 2003; Nelson, Bolia, Ericson and McKinley, 1998).

It is important to note that in many of these studies the enhancement in intelligibility and overall performance is not trivial. Brungart, Ericson and Simpson (2002) examined the effect of spatial separation on performance, with one, two, or three same-sex interfering talkers during an auditory monitoring task. They found that with one interfering talker spatial separation increased performance by around 25%. In the case with two or three interfering talkers, spatial separation of the talkers nearly doubled the percentage of correct responses (See Figure 9).

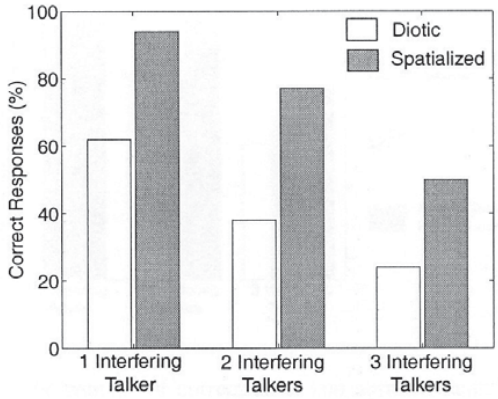


Figure 9: Comparison between diotic and spatialized listening conditions as a function of the number of interfering talkers (Reproduced from Brungart, Ericson and Simpson, 2002)

Although this substantial body of research is compelling in terms of demonstrating the potential advantages of using spatial audio displays to segregate concurrent communications messages, most of this work has been carried out primarily in highly controlled laboratory settings using single task paradigms. An important step in the transition of spatial auditory technology to real-world applications is the demonstration of its benefits in more realistic and complex task environments. There have been a small number of papers that have attempted to do this. Nelson and Bolia (2003) went some way towards testing the usefulness of spatial auditory displays in a simulated Airborne Warning and Control System (AWACS). They manipulated listening conditions (spatial versus non-spatial), chatter level (higher versus low) and mission phase (ingress, retargeting, and egress) and asked aircrew to monitor short phrases listening for a call sign and responding by identifying the appropriate colour-number combination that followed their specific call sign. This task was conducted throughout a Close Air Support mission where aircrew were required to track a package of fighter aircraft. Results indicated that speech intelligibility was degraded during the most difficult mission segments, and that spatial audio alleviated the degree of this degradation. In addition, spatial audio was associated with faster response times for correct identification of critical call signs. Post-experimental surveys indicated that operators rated spatial audio technology as valuable for improving communications effectiveness.

Haas, Gainer, Wightman, Couch and Shilling (1997) examined how accurately helicopter pilots could process radio communications information in a simulated cockpit environment when the messages were presented under different listening conditions (diotic [two messages both played to each ear], dichotic [one message played to the left ear and a different message played to the right ear] and 3-D audio). Helicopter pilots participated in a 30-minute flight scenario where they were required to perform target acquisition tasks and respond to various malfunctions during each simulation flight. During the flights each pilot was asked to monitor the simultaneous radio transmissions on three radio channels and to transmit a reply radio message when they heard their call

sign. Pilots obtained the greatest number of correct message identifications when using 3-D audio, less with dichotic presentations and the smallest accuracy for diotic presentations. This was a relatively difficult radio communications task involving the perceptual separation of three simultaneous messages.

In a simulated air traffic control task MacDonald, Balakrishnan, Orosz and Karplus (2002) measured improvements in the detection of auditory warnings when using virtual 3-D audio. The simulation was designed to reproduce some basic elements of flight test control operations, with participants being asked to monitor progress during a flight test where pilots have a schedule of manoeuvers to perform. During the flight test schedule the test controller was required to simultaneously monitor a number of communications and data streams. Participants were instructed to select the detected warning from a list of warnings, e.g. " low altitude alert" (maintained in a graphic display area), and then identify the location (speaker) of the warning. Results indicated a significant effect of configuration when the four simultaneous sounds were separated into at least two static locations, compared to when sounds were generated from only one location.

Finally, Brungart, Ericson and Simpson (2002) note that there are techniques other than spatial separation that can be used to improve intelligibility in multitalker speech displays. These include improving the signal-to-noise ratio, reducing the number of competing talkers, changing the voice characteristics of the talkers and altering the relative levels of the talkers. They conclude however that the most effective and efficient way to improve the usefulness of multitalker displays is to spatially separate the locations of the competing talkers. They point out that one of the biggest advantages of using spatial separation is that it improves the intelligibility of all the talkers in an approximately equal manner, while the other techniques tend to increase the intelligibility of only one of the competing talkers.

4.3 Air traffic collision avoidance

A number of studies have evaluated the usefulness of 3-D audio for Traffic alert and Collision Avoidance Systems (TCAS). Begault and Pittman (1996) compared a visual TCAS condition against a condition where TCAS information was supplied via 3-D audio without a visual display. Aircrew were required to fly a commercial airline simulator and during the flight were given a number of traffic warnings and were asked to verbally indicate when they had visually acquired the aircraft associated with the warning. They found that acquisition times were faster by 19% for the 3-D audio condition, and the number of targets acquired was higher. In a further study using a similar experimental design Begault, Wenzel and Lathrop (1997) compared a visual TCAS condition against a 3-D audio condition paired with a visual display, which differed from the previous study where no visual display was presented with the 3-D audio. They once again found a reduction in acquisition time of 19% for the 3-D audio condition, although unlike Begault and Pittman (1996), there was no difference in the number of targets acquired.

Oving and Bronkhurst (1999) compared performance under four conditions including monophonic sound (no 3-D audio) without verbal information, monophonic sound with

verbal information, 3-D audio without verbal information, and 3-D audio with verbal information. The verbal information consisted of speech that designated the quadrant in which a target was located (e.g. "Traffic up left"). Aircrew were given traffic alerts during a simulation flight that involved aircraft approaching an airfield and landing. Oving and Bronkhurst reported that the average response time for the 3-D audio conditions was 20% faster than that for the monophonic conditions, and that performance in conditions where verbal information was present was better than that in conditions where verbal information was absent.

4.4 Cockpit warnings

Haas (1998) evaluated the impact of 3-D audio on response times to the presentation of warnings during a helicopter simulation trial. Pilots were asked to respond as quickly as possible when warning functions occurred. The warnings included fire in left engine, fire in right engine, chips in transmission and shaft-driven compressor failure. The warnings were presented in three different modes including visual only, visual plus 3-D audio speech, visual plus 3-D audio non-speech signals. Haas found that response times were significantly faster in the two conditions in which 3-D audio was provided (30% faster for the visual plus 3-D audio speech condition than in the visual only condition and 26% faster for visual plus 3-D audio non-speech condition than in the visual only condition).

4.5 Spatial Disorientation Countermeasure

Bles (2004) reports that over the last decade spatial disorientation has been identified as a major cause of flight accidents. He cites a number of potential causes including increasing manoeuvrability of fighter aircraft and helicopters, increased pilot workload, increased demands of flight conditions (i.e. more night flying, more headdown flying, more low level flying), and the use of display systems that introduce additional visual frames of reference (e.g. off-boresight targeting). Bles also lists a number of countermeasures for spatial disorientation including the use of 3-D audio. One of the initial studies to use acoustic information to aid spatial orientation was Lyons, Gillingham, Teas, Ercoline, and Oakley (1990). They evaluated the impact of an acoustic orientation display that provided information on airspeed (represented using repetition rate of the auditory signal with increases in speed associated with increases in repetition rate), vertical velocity (represented using amplitude modulation rate with increases in velocity associated with increased pitch) and bank angle (represented using right/left lateralization with a louder signal on the side of the head that was in the same direction as the bank). Lyons et al (1990) reported that the acoustic display was a useful indicator of aircraft orientation in the absence of visual cues with the bank angle being the most effective.

Bles (2004) concludes that acoustic displays could play a role in avoiding spatial disorientation where pilots could match the auditory spatial-orientation information with the environment and any "mismatch between the auditory information and the pilot's natural orientation percept will trigger the pilot to restore correct orientation" (Bles, 2004, p. 526). As one example, 3-D audio could be used to indicate the gravity vector. Bles states that at the very least, 3-D auditory displays could reduce the pilot's visual workload, allowing more time to be spent of primary mission tasks, thereby reducing the chance of

spatial disorientation occurring. This role for 3-D audio however needs more research to determine whether such potential can indeed be realised.

5. Auditory research at DSTO

For the past ten years DSTO has been engaged in an extended program of research on a range of topics associated with the use of auditory warnings in military aviation cockpits. These topics include not only 3D audio, but also more general topics as well, including choice of type of auditory warning. In all areas the overall design philosophy has been to give back to operators in the cockpit the everyday ability they have to extract information about their world from the auditory inputs they receive. In the real world individuals can tell a lot about what is going on around them based on what they hear. When they hear a door shut, they can generally identify some key characteristics of the door (such as whether it is a fly-wire screen or a more solid wood door, whether it was slammed in anger or not) as well as a clear idea of where the door is. In the everyday world this information complements what the visual system tells us about what is occurring, and the two systems work in collaboration to provide as detailed a picture as possible. For example, the auditory system alerts individuals as to what is occurring in areas where they are not looking. All of these everyday opportunities are not currently provided to operators in the cockpit. The aim of auditory research at DSTO is to redress this discrepancy.

5.1 3-D audio

The initial focus of the 3-D audio research programme was technical, examining methods and techniques associated with the generation of high-fidelity 3-D audio. Recording techniques, HRTF generation, convolution algorithms and hardware selection were all examined in an attempt to generate accurate 3-D audio.

To date, the technical program has resulted in an ability to generate 3-D audio that provides localization accuracy equivalent to listening to sounds in the real-world (see Martin, McAnally and Senova, 2001 in Section 3.2 and Figure 5). Technical work is continuing as improvements are identified to simplify procedures and requirements and to generate a more robust methodology suitable for use in military aviation environments.

In addition to examining methods and techniques for the generation of 3-D audio some research programs have been directed at examining issues associated with the implementation of 3-D audio. For example, there has been a research program aimed at the refinement of interpolation techniques to enable smoother movement and to reduce the number of locations required during the recording process. In the last few years, the emphasis has shifted further to the examination of issues associated with the application of 3-D audio (i.e. how best to use 3-D audio). This has involved the systematic examination of the use of 3-D audio during different tasks.

The following is a list of topics that have either been investigated in the past, are currently under investigation, or have research activities planned in the near future:

- Free field versus virtual (3-D audio) spatial hearing
- Spatial fidelity of 3-D audio
- Variability in the headphone-to-ear-canal transfer function
- Localization of virtual sound as a function of head related impulse response duration
- Interpolation algorithms
- Impact of hypoxia on the use of 3-D audio
- Impact of G forces on the use of 3-D audio
- Effect of signal content on 3-D audio
- Impact of head movement on the use of 3-D audio
- Aurally guided visual search
- Effects of supplementing Head-Down Displays with 3-D audio during visual target acquisition
- The use of virtual audio for the spatial segregation of competing speech
- Segregation of multiple talkers in the vertical plane
- Impact of visual movement on accuracy of 3-D audio systems
- Use of multiple sounds in 3-D audio systems
- Spatial disorientation countermeasure

The contact details of those researchers responsible for the auditory research program at AOD/DSTO are included in Appendix A and a more detailed list of publications is included in Appendix B.

5.2 Auditory warnings

In addition to 3-D audio there are other areas of auditory research being conducted at DSTO which includes research regarding more general design issues associated with the use of auditory warnings. There are three key criteria to be addressed when designing auditory warnings (Stanton and Edworthy, 1999). First, auditory warnings should alert the operator to the fact that there is a situation that requires attention. Second, auditory warnings should provide the operator with information about the nature or identity of the situation requiring attention. Third, auditory warnings should guide the user towards the appropriate course of action to deal with the situation. 3-D audio helps address some of these criteria but other properties of auditory warnings also need to be considered before a warning can be considered completely successful. For example, the advantages of providing location information using 3-D audio techniques can be eroded if the warning fails to alert the operator and capture his/her attention. With this in mind, in addition to research related to 3-D audio, other work has been conducted at AOD, DSTO that examines aspects of auditory warnings that also need to be considered.

Auditory warnings can be divided into two distinct categories, verbal and non-verbal. Both have their relative advantages and disadvantages. Verbal (speech) signals have the advantage that negligible learning is required, and are also capable of conveying complex

information, although because of the serial nature of speech it may take some time to completely process such information. A disadvantage for speech however is that in addition to masking by background noise (which all types of auditory warning are subject to), speech is also subject to informational masking by other speech. This is particularly true in the aviation environment where communication between crew members, and communication over radios are such an important part of the operational environment.

Abstract sounds (e.g. simple tones) are commonly used as a non-verbal alternative in warning systems. They can convey information quickly, and they may not be as susceptible to information masking in speech-rich environments, but it is difficult to learn large sets of abstract sounds and they can only convey a limited amount of information. Auditory icons (environmental sounds) are another type of non-verbal sound that has potential for use as auditory warnings. This is based on the observation that in everyday life we manage to extract information about our environment through the identification of naturally occurring sounds. One possibility is that we can increase the effectiveness of aircrews' responses to auditory warnings by exploiting this pre-existing ability. Auditory icons may share with abstract sounds a resistance to informational masking and are also capable of conveying information quickly. They have an added advantage though, compared to abstract sounds, in that they have the potential to convey more information and that they are easier to learn.

AOD, DSTO has been involved in exploring the potential of auditory icons to be used as another type of auditory warning. In one of a number of studies, Stephan, Smith, Parker, Martin and McAnally (2000) have compared the learning and retention of auditory icons compared to speech and abstract sounds and found that auditory icons were learned and retained as easily as speech, and that both auditory icons and speech performed better than abstract sounds. Further research is being conducted to determine whether icons have any advantage over speech for tasks conducted in the presence of high workload. Techniques are also being explored that will allow a single auditory icon to convey more than one piece of information about an important event e.g. in addition to providing information regarding the identity of the event, information regarding size, distance and motion associated with the event (see Stevens, Brennan and Parker, 2004 for a more detailed description of this research).

5.3 Research capabilities at DSTO

5.3.1 3-D Audio Laboratory

To support the research activities described in Section 5.1 a dedicated 3-D audio laboratory was designed and installed at AOD, DSTO. The laboratory consists of an acoustically treated room (3m X 3m) lined with sound absorption materials designed to reduce both external noise and sound reflections (background noise levels within the room < 10 dB SPL and an absorption coefficient ≥ 0.99 down to 800 Hz). Inside the room is a 1m-radius hoop suspended from the ceiling on which a loudspeaker is mounted (see Figure 10). The mounting mechanism is designed such that the speaker can be rotated 360° in the horizontal plane and moved in the vertical plane over a range that extends from 50° below

the level of the ear canals to 80° above. Movement of the speakers is achieved using stepping motors that are computer-controlled. At the centre of the hoop is located a swivelling chair where participants are seated during experimental trials. An acoustically transparent cloth sphere that is supported by thin fibreglass rods obscures the participant's view of the hoop and loudspeaker. The inside of the sphere is dimly lit to allow visual orientation.

Listener's wear a headband upon which is mounted a magnetic-tracker receiver and a laser pointer. Following the presentation of a sound participants are asked to indicate the location of the sound by positioning the laser pointer to point to a place on the surface of the cloth sphere that is coincident with the apparent location of the sound. Once the laser has been positioned, participants press a handheld button and the magnetic-tracker coordinates are recorded and used to calculate the response location.

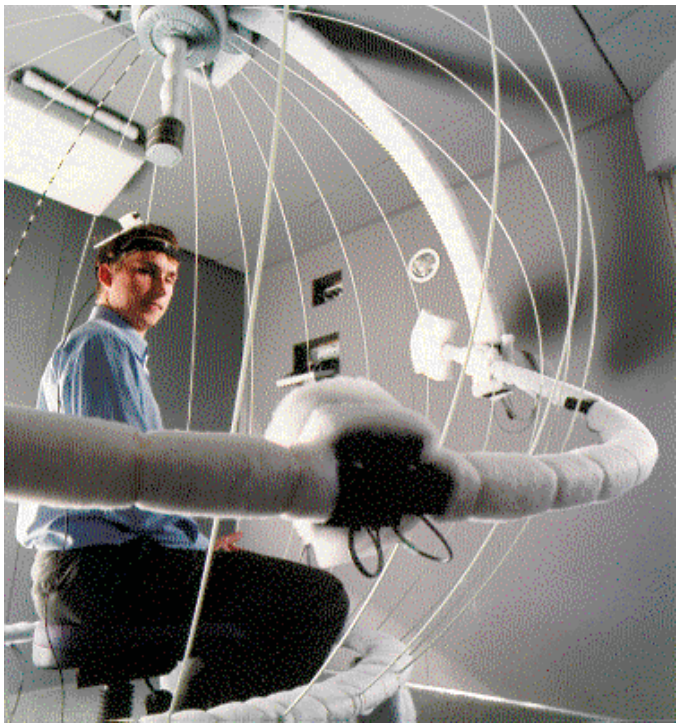


Figure 10: 3D audio laboratory at AOD, DSTO featuring the hoop with speaker and listener wearing headband with head tracker and laser pointer. During experiments a cloth is draped over the fibreglass rods so that the listener cannot see the location of the speaker

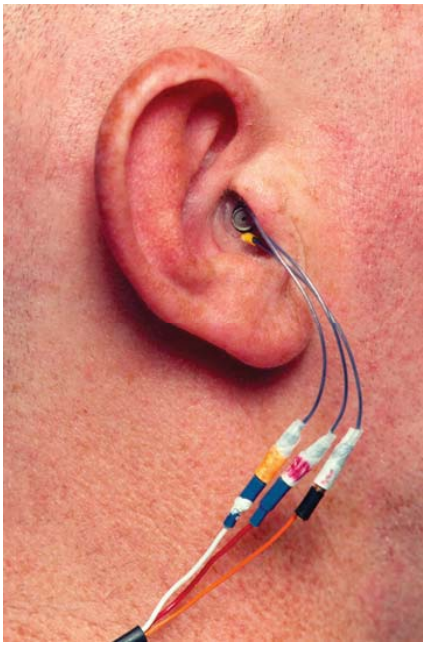


Figure 11: Photograph showing the placement of the microphone during the application of the blocked-ear-canal measurement technique

To generate a 3-D audio display a set of HRTFs are constructed for each participant. These are generated using a "blocked-ear-canal" measurement technique (see Martin, McAnally and Senova, 2001) where miniature microphones are placed in the participant's left and right ear canals such that the diaphragm is at least 1mm inside the entrance of each ear canal (see Figure 11).

In addition to, and associated with the 3-D audio laboratory, is a general capability to generate, edit, and mix sounds using a variety of software systems. AOD, DSTO uses Tucker-Davis hardware with customised software designed and constructed on site (see Section 7.3.3).

5.3.2 Air Operations Simulation Centre

The simulation trials used to test the application of 3-D audio were conducted at the Air Operations Simulation Centre (AOSC). The AOSC provides a distributed, real-time, human-in-the-loop simulation environment for the conduct of research into the performance of aircrew and aircraft systems in operational scenarios. The AOSC is a flexible, interactive and reconfigurable simulation facility incorporating a range of components from high to low fidelity. These include a Partial Dome Display System, Cube Display System, advanced image generation systems, Helmet Mounted displays, fixed and rotary wing cockpits, scenario generation and tactical environment software, tools for the generation of visual databases and avionic instrumentation panel displays.

A simulation within the AOSC is formed by running a collection of software modules that are interconnected and synchronised by scheduling software developed within the AOSC. Typical software modules required for a simulation conducted in AOSC include flight models, cockpit interface, flight instruments, image generation, experimental control software, data logger, communications and external connectivity (See Figure 12). Part of the integration of a new module into the AOSC includes defining and registering all inputs and outputs to and from the module. The Distributed Architecture Memory Manager then stores all inputs and outputs from all software modules and links input and outputs between modules where required.

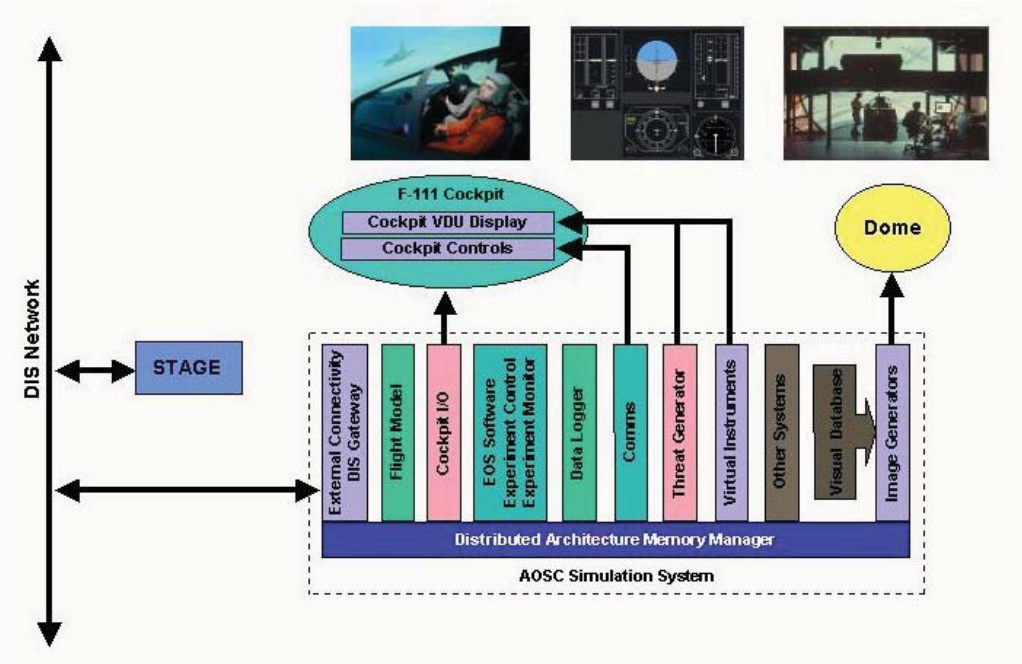


Figure 12: Diagram of an example of the software module configuration used in the Air Operations Simulation Centre (AOSC) during a simulation trial.

The AOSC software systems are hosted on a mixture of hardware including Silicon Graphics mainframes and clusters of PCs. The out-of-the-window view is produced by an image generator (Vega 3.2) and can be projected via a range of capabilities that include high-resolution and wide field-of-view options (Partial Dome display) to lower fidelity systems with smaller fields of view (3 channel, flat screen low cost systems).

AOSC has a range of cockpit modules including fast jet, transport and rotary wing cockpits, used during simulation trials as required (see Figure 13 for an example). In most cases the cockpit consists of all external surfaces and structures (with the exception of the canopies) to accurately replicate the viewing conditions for aircrew seated in the cockpit. The cockpits generally contain flight controls and modified instrument panels to provide flight information; navigation and communication systems and other systems (e.g. weapons systems) are added if required for the trial being conducted.

Figure 13: Photograph of an example of a cockpit module. This photograph shows an F-111C cockpit where the main instrument panel has been replaced by three computer monitors which allow for the display of software versions of primary and secondary flight instruments, as well as other types of displays.

In addition to the capabilities required to simulate a flight mission the AOSC has a range of capabilities for recording data during the conduct of simulation trials which allow for detailed testing of new systems (e.g. 3-D audio) to evaluate their potential usefulness in aviation environments in the presence of workload and under conditions that approximate the real work environment.

6. Requirements for a 3-D audio system

As for any system, the requirements for a 3-D audio system are dependant on the nature of the proposed application of the system. As mentioned in Section 4, 3-D audio can be used for a range of potential applications. Some of these applications have more demanding perceptual requirements than others. For example, the identification of the location of an incoming missile has a requirement for absolute accuracy (e.g. no front-back errors), high resolution in terms of location information (possibly up to 1° accuracy), and an ability to generate the perceived location of an event in the absence of a corresponding visual image. These requirements are very different to the requirements of an application that is attempting to segregate simultaneous communication inputs. In the latter application the specific location of each communication input may be less important than the ability of the listener to separate the simultaneous inputs, therefore, lower resolution in terms of location information is required and absolute accuracy is less important. These different requirements will have varying degrees of impact on the final design of a 3-D audio system, including both hardware and software components. It is possible however, to outline a general set of requirements for a 3-D audio system, for although the eventual

application of the system will drive the detail of any specific design, most applications share a number of common features.

The process of generating 3-D audio has a number of distinct stages. The method of implementation for each of these stages can vary substantially, but the functionality of each stage must be achieved for the accurate implementation of 3-D audio. These stages include identifying the location of the event of interest, spatialization and output.

The first stage involves the identification of the location of the event to be displayed. For example, it might be the location of an incoming missile if the application for the 3-D audio was to be a part of an Electronic Warfare system on an aircraft. To determine a unique location for the event to be spatialized requires the calculation of location by taking into account a range of relevant variables such as aircraft orientation, the location of the event in the world and the listener's head position. In the second stage, once a unique location has been identified the appropriate cues for the generation of 3-D audio need to be reproduced and embedded in the sound to be spatialized. Finally in the last stage, the 3-D audio sound needs to be output to the listener. The functional components of each of these stages are outlined below, and the high-level hardware and software requirements to achieve this functionality are discussed.

6.1 Functional requirements

The functional requirements for each stage are outlined in Figure 14, which describes the general requirements for a system designed to deliver a single sound to a single listener. Whilst some of these functions are represented separately it is possible to combine some of these functions within a single process.

6.1.1 Stage 1: Event location identification

The first functional stage is to identify the location of the event to be spatialized relative to the centre of the head of the listener (i.e. to calculate location in coordinates relative to the listener). This requires identifying the location of an event (e.g. a threat location as identified by a Radar Warning Receiver) and compensating for the listener's head position and orientation. In an aircraft the location of the event may also need to take into account aircraft orientation. This would require additional inputs identifying aircraft position and orientation so that such information could be factored into the computation of the final calculations.

For some applications aircraft location and orientation are already factored in and therefore the only compensation required is an adjustment for head position. For example, many Electronic Warfare systems (such as a Radar Warning Receiver) calculate threat location relative to the aircraft and compensate for aircraft orientation to adjust the signal processing of inputs from receivers at different aircraft tilt angles. Other systems may provide location information, but in different coordinate systems. For example, systems may provide location in earth coordinates (rather than in aircraft coordinates). Some translation may therefore be required to determine aircraft heading and orientation in order to derive location relative to the aircraft rather than independent of the aircraft.

Having calculated the location of an event in coordinates relative to the listener, this information needs to be continuously updated to compensate for head movements, otherwise the location of the sound will move when the head moves. For example, a stationary external event located at 90 degrees to the left of the aircraft should not move relative to the aircraft as the listener rotates his/her head to the right. Without compensation for head movement the sound will appear to move with the head as it is rotated because the system will continue to project the location of the external event so that it remains 90 degrees to the left of the head. For example, if the listener were to make a 90-degree head rotation to the right, this would result in an apparent location of the sound at 0 degrees relative to the aircraft. In this situation the location of the external event has not changed, rather the position and orientation of the listener's head has changed. By compensating for head position the event location will be updated so that it remains constant at 90 degrees to the left of the aircraft regardless of head position and orientation.

The module identifying event location may have multiple inputs from different sources (i.e. different systems). For example, an Electronic Warfare system may have a number of potential sensor systems. Event Source 1 may be a Radar Warning Receiver, Event Source 2 a Missile Warning Receiver and Event Source 3 a Laser Warning Receiver. The output from the selection module is a single location for each event, that will deliver a virtual spatial sound at a position correlated with the actual location of the event in the external world. In the simple example in Figure 14 only one event location can be delivered at any single moment in time, although in an appropriately designed system it is possible to output multiple simultaneous sounds (the requirements associated with the generation of a multiple 3-D audio sounds will be considered later in Section 6.4).

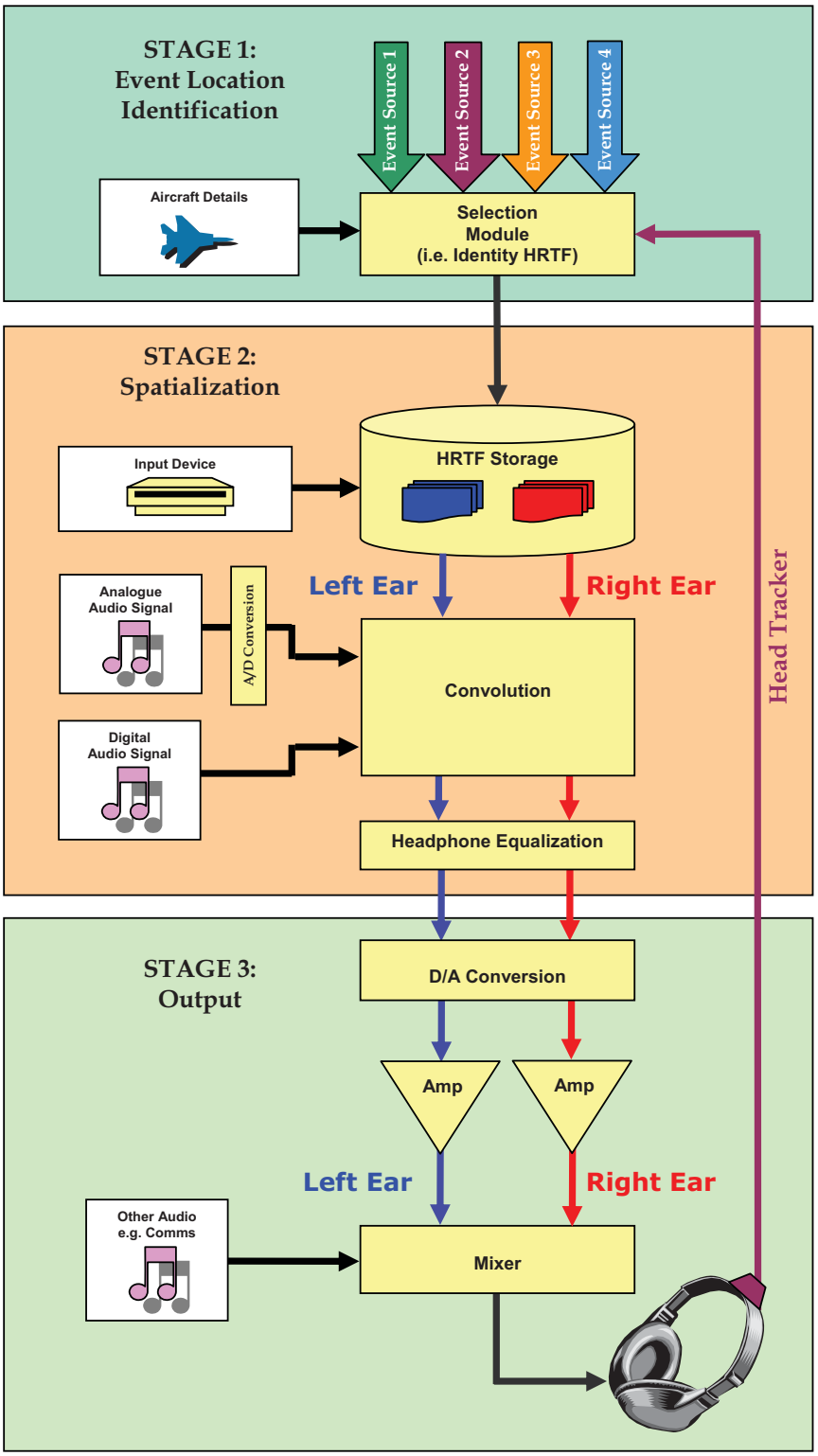


Figure 14: Representation of the general functional requirements for a 3-D audio system

6.1.2 Stage 2: Spatialization

The second functional stage involves using the output coordinates from the selection module (i.e. Stage 1) to choose a pair of filters from the complete set of possible locations which are stored somewhere in the 3-D audio system. If the coordinates from the selection module do not have a corresponding filter pair then the most optimal response would be to use an appropriate interpolation algorithm to generate a filter pair for those coordinates. Interpolation is the process whereby an intermediate location is calculated based on derivation from neighbouring positions. Many different interpolation algorithms are available ranging from simple linear interpolation to more involved weighting algorithms and complex spline formulae.

The interpolation may be calculated online (real time) as required, or offline where a large number of locations have been calculated and then stored in the system ready for use when required. Interpolation becomes necessary when there is a requirement for high-resolution, in terms of the number of filter pairs used, to provide smooth apparent motion as a sound moves. A large spatial separation between virtual locations results in the sound appearing to jump from one location to another, instead of sliding smoothly across smaller steps. It is worth noting that the better the interpolation algorithm, the fewer recording locations that are required. The use of real time interpolation algorithms reduces the requirement to store large numbers of filter pairs. The use of stored offline interpolation can result in the storage of large numbers of filter pairs (some systems store a filter pair for every location at intervals of one degree), so that a filter pair is available for all possible locations. The ready availability of storage options makes offline interpolation viable, with the added benefit that no interpolation calculations are required, reducing demands for processing capability and reducing the likelihood of any latency.

Following the selection of the appropriate filter pair, these filters need to be applied to the audio signal that is to be spatialized. This process occurs in the digital domain, therefore if the input audio signal is in an analog form it will need to be converted to a digital form through an Analog to Digital (A/D) conversion process. The application of the filter pairs to the input audio signal is achieved via the process of convolution. Convolution occurs in the time domain and is a time-indexed multiplication and summation operation performed on two numerical arrays (hence the need to transform the signal into a digital format), one array being the filters the other array being the audio signal. The output from this convolution process will be a sound with all of the relevant spatial cues embedded into the audio signal.

The last function in this second stage involves making some adjustments to the audio signal to remove any artefacts caused by the response characteristics of the transducers in the headphones. Headphone transducers do not as a rule faithfully reproduce every aspect of an audio signal. The changes produced by the headphones need to be removed from the signal so that the listener receives a sound at the eardrum that contains only the correct spatial cues required for accurate localization. This is achieved by measuring the characteristics of the headphones used by the listener and then applying a filter that produces the inverse of any changes caused by the headphones to negate these characteristics. While it is important that headphone equalisation occurs, it is possible to

achieve this during the convolution process when the location filter pairs are applied to the input signal. This therefore represents an example where the functionality is required but the exact method of application could vary substantially.

6.1.3 Stage 3: Output

The final stage following the application of the filter pairs and headphone equalisation (i.e. Stage 2) is to prepare the audio signal for output to the listener. Both the convolution and headphone equalisation processes will have occurred in the digital domain. Most headphones however require an analog signal. Therefore prior to delivery to the listener the audio signal needs to be converted via a Digital to Analog (D/A) process into the analog domain. The signal may also require some amplification before input to a mixer, where the signal is combined with other non-spatialized audio inputs (e.g. radio communications) if present, before delivery to the listener via headphones.

6.2 Hardware requirements

Each of the functional stages has some specific hardware requirements. Stage 1 (Event location identification), which includes the selection module, will require some form of processing power and some form of interface to the various inputs which provide the basic data needed to calculate the location coordinates for each event to be spatialized. The source of these data will vary from platform to platform. Modern platforms will have a databus and a mission computer, which will provide reasonable access to the data required. Older platforms may have neither databus nor mission computer, and separate federated interfaces may be required to systems such as navigation aids (e.g. GPS).

Stage 2 (Spatialization), which adds the spatial information to the audio signal, has the most hardware requirements. Some form of memory is required to store the HRTF filter pairs. The size of this storage capability depends on the location resolution required, and various parameters of the HRTFs such as sample rate, filter length, etc. Some form of input device is required to load the HRTF filter pairs into the storage capability. The input device is required regardless of whether individualized or non-individualized HRTFs are used because the presence of some kind of HRTF filter set is necessary for the production of 3-D audio. What may vary is the frequency of use of the input device. In the case of non-individualized HRTFs the input device would be used infrequently and only when changes are made to the single HRTF set used in the system. In the case of individualised HRTFs the input device would be required every time the listener changes.

Depending on the origin of the audio signals, there is a requirement either for an interface to allow the input of the audio signal if it is generated by other systems external to the 3-D audio system, or for more storage if the signals are generated within the 3-D audio system. There will be a requirement for Analog to Digital (A/D) conversion processor if the audio signals are in an analog format. Some additional storage capacity may also be required for the headphone filters and processing capacity will be required to conduct the convolution and headphone equalization processes. It should be noted that the provision of processing capacity for the convolution process is one of the most important hardware requirements

for the 3-D audio system since this represents one of the most critical functions of the system.

In Stage 3 (Output), which is involved in the delivery of the processed signal to the operator, a D/A converter is required as well as a potential requirement for some amplification and mixing capability. Within a cockpit environment this may result in changes to the communications system. The communications system in most existing cockpits is usually low in fidelity, consisting of a monophonic system with a low-pass frequency response similar to that of the household telephone, neither of which are suitable for a 3-D audio system. The signal must be played on transducers (headphones) that are stereo, and if elevation and front/back discrimination is important the transducers must be broadband in their frequency response (i.e. must include the capability to reproduce high frequencies). Again the difficulty here is that depending on the airframe in question, there may be changes required to existing aircrew helmets and ICS systems. Some capability to monitor and measure the position and orientation of the head is also required in order to produce sounds that are geo-referenced in space. In some cases such a capability may already be resident within the cockpit, for example when a helmet-mounted display is fitted where the monitoring of head position is also a requirement for the helmet-mounted display system. In cases where such a capability does not exist, some form of head tracker is required. There are different types of head tracker available including optical, electromagnetic, ultrasonic and miniature inertial guidance systems.

The hardware requirements for Stage 3 may include changes to the cockpit involving safety-critical systems which might involve re-certification. Most of the hardware components are relatively simple and readily available in various commercial forms. The number and type of hardware components can vary substantially as a function of the specific design. For example, the A/D and D/A conversions can be achieved using specialised processors designed specifically for these types of conversions. Alternatively, it may be possible to conduct these conversions using specialised software and general-purpose processors, although there would be some cost in terms of speed and accuracy. A final solution will most likely be dependent upon the number of conversions required and the desired speed of the calculations associated with these conversions. A point worth noting is that the entire system must operate in real-time with small delays and fast update rates. Such requirements may necessitate the choice of specialised processors.

6.3 Software requirements

In addition to hardware requirements there are a number of requirements related to the construction of software to drive and utilise the hardware fitted. In contrast to the hardware components, some of the requirements for this software are quite demanding, as it is the software that makes the greatest contribution to the overall capability of the 3-D audio system. In addition, there are a large number of potential solutions, and care must be taken to identify the most optimal of these.

In the first stage there are software requirements for a module that will collate the relevant data required to allow the calculation of a single location for each event. Having calculated the location, the module is then in a position to identify the most relevant HRTF filter pair.

In the second stage the software is required to retrieve the relevant HRTF pair, to initiate the conversion of the input audio signal into a digital form, and to implement the convolution process and headphone equalisation. In the final stage the software is required to transform the filtered audio signal back into the analog domain and ensure it is amplified and passed through the mixer before delivery to the listener. The exact role of the software in this stage will depend on the design solution. If specialised hardware processors are part of the final design solution, then the role of the software becomes one of process control. If specialised hardware is not part of the final design solution then the software involves computation as well as process control.

It is worth noting that apart from the specific functional requirements that are directed at driving the hardware involved and conducting the signal processing that needs to occur, there are a set of broader requirements that need to be considered during the design of the software. These include the overall principles of modular design and flexibility that need to be applied to the software architecture. Because 3-D audio is a new technology it is likely that as our understanding of the use and application of the technology increases, improvements in certain aspects of the system will become apparent, and a software design that will accommodate change more readily would be advantageous. Areas that may well change are any components that apply algorithms (e.g. online interpolation) where improvements to the algorithms could conceivably occur. Also, there may be additions to inputs into the 3-D audio system. For example, if the application associated with the 3-D audio was as part of an Electronic Warfare system then additional sensor systems might be added.

One component that will definitely require some degree of flexibility is any section associated with the input, storage and application of HRTFs. The HRTFs can be recorded at different locations, using different coordinate sets and have different sampling characteristics. It is important that the software be capable of accepting HRTFs in a variety of formats.

6.4 Multiple sources and multiple users

The system outlined in Section 6.1 describes the requirements for a 3-D system that will allow the generation of a single sound to a single listener. As such it represents the simplest form of a 3-D audio system. It is more likely that the fielded system will have the additional requirements of being able to generate multiple sounds simultaneously and that there may be more than one listener involved. Such additional requirements will not change either the number of functional stages or the nature of the stages. More than one sound source or listener requires some degree of duplication within the 3-D audio system. The primary component involved being in Stage 2 where spatialization of the auditory signal occurs. This can be achieved either by duplication in the hardware with a separate hardware processor for each sound source and/or each listener, or duplication in the software with the convolution processing being shared across the available hardware processors.

The limitation on the number of sounds required is going to be determined not so much by hardware and software limits, as by the perceptual limitations of the listener (i.e. how many sounds can the operator use effectively before the workload increases and becomes unmanageable). Once again, this number is likely to vary as a function of the nature of the application and the symbology used. The number is likely to be less when the operator is required to track and interpret each individual sound, as opposed to applications such as communications segregation where the primary function of the 3-D audio system is to keep simultaneous signals separate so that they can be recognized at all times. Regardless of the application, the number of sounds required is likely to be relatively small (e.g. 4-8 sounds), rather than large.

6.5 HRTF generation

It is worth recognizing that whilst the method of generating HRTFs has minimal impact on the design of a 3-D audio system (other than to ensure the system can input and use the HRTFs), the overall performance of the 3-D audio system is to a large extent dependent on the quality of the HRTFs. It is important therefore that this part of the process be kept in mind, and that appropriate steps are made to ensure that whatever the method of implementation chosen, the validity and accuracy of the HRTFs are optimized.

7. Existing Commercial 3-D Audio systems

At any given time there are a large number of products on the market which all claim to provide spatial (3-D) audio of some type or other. It is not the intention here to review all the available products, but rather to review a subset of these products to provide examples of the various types of commercial 3-D audio systems that are available at present. It should be noted that these products change rapidly, and new products arrive regularly, so any review will only reflect the state of the information available of the time of writing.

The majority of 3-D audio solutions are essentially a software interface to hardware systems capable of producing audio. The hardware in question is generally of the form of PC soundcards, which explains the fact that many of the companies producing 3-D audio software are the same companies producing sound cards for the PC market. The function of this hardware is to facilitate the production of audio in a similar way that graphics cards are used to facilitate graphics processing on any personal computer. However, the similarity of these subsystems ends here. As graphics technology development moves continually toward dedicated hardware support for 3-D graphics (as opposed to the less computational demanding 2-D graphics requirements), audio technology is not following this trend yet. Audio technology development is primarily driven by increases in audio fidelity, which are developments consistent with the requirements for high-quality 3-D audio. However, dedicated hardware support for 3-D audio in this technology market doesn't share the same degree of attention as the graphics subsystems. The growth in general digital signal processing (DSP) technology in soundcards is often mistakenly understood as dedicated hardware support for 3-D audio. As it stands, 3-D audio solutions on the PC are still predominately software based, calling upon general DSP

hardware for computations. One of the differences between software options is the degree to which the software is dedicated towards a specific hardware solution. It should always be remembered that commercial 3-D audio has severe design constraints as a rule. The target hardware has to be low-cost, readily accessible, and the function of the 3-D audio is to supplement the visual experience, mostly during games. As a supplementary experience then, the initial emphasis of 3-D audio was in terms of environmental acoustics in an attempt to render realistic auditory environments. As the game player moves from environment to environment, they not only see but also hear the old west, a space station, underground tunnels, cathedrals, etc.

Within the commercial market the term '3-D audio' and 'spatial audio' are used interchangeably and are used to cover a wide and confusing spectrum of technologies. This confusion can be clarified by sub-dividing the technologies into three key categories, each offering increasing levels of sophistication and capability.

The first category is Extended Stereo (also known as stereo expansion). This technology processes a 2 channel audio stream (stereo) to add 'spaciousness' and make the sounds appear to originate from locations well beyond the actual left/right speaker locations.

The second category is Surround Sound. This category encompasses all technologies that create larger-than-stereo sound fields through multi-channel audio streams with multiple arrays' of speakers (i.e. >2). The most common examples of surround sound are Dolby ProLogic or Dolby Digital with a conventional 5.1 speaker set-up (5 wideband speakers plus a low-frequency sub-woofer). This category also includes virtual surround technologies that aim to reproduce the multi-speaker set-up (typically 5.1) through virtual 3-D audio techniques (i.e. over headphones).

The third category is 'True' positional 3-D Audio. This category includes the technology capable of positioning sounds anywhere in the three-dimensional space surrounding the listener, giving the best approximation of the real-life listening experience.

7.1 PC Based 3-D Audio Solutions

7.1.1 Interactive Audio Special Interest Group - 3D Audio Workgroup

The Interactive Audio Special Interest Group (IA-SIG) was formed with the purpose of engaging developers of audio software, hardware, and content to freely exchange ideas in order to improve the performance of interactive applications by influencing hardware and software design, as well as leveraging the combined skills of the audio community to make better tools. The 3D Audio workgroup focuses on creating 3D Audio rendering guidelines to define more realistic audio environments. This effort has lead to extensions to the Microsoft DS3D API (see Section 7.1.2) to enable hardware acceleration, and to the publication in 1998 of the 'IA-SIG Interactive 3D Audio Rendering and Evaluation Guidelines (Level 1)' describing "minimal acceptable" 3D audio features for all platforms. This guide has set the standard to which many developers now adhere.

The requirements for certification under this guide are:

- Playback of 8 simultaneous sources
- Minimum sample rate of 22050Hz, 16-bit output
- Object and listener 3D position (x, y, z)
- Object and listener velocity
- Listener orientation
- Object orientation and radiation pattern
- Distance effects rendering
- Doppler effects rendering
- True 3-D positions (x,y,z) effects rendering
- Radiation model effects rendering
- All rendering in real-time without audible artefacts or latency

7.1.2 Microsoft DirectSound3D (DS3D)

Built into Microsoft Windows are a suite of multimedia Application Programming Interfaces (APIs) called Microsoft DirectX, and Microsoft offers 3-D sound as part of the DirectSound module. While the predominant focus of the DirectX suite is its graphics capabilities, the DirectSound module includes DirectSound3D (DS3D) to act as a 3-D audio API. If a soundcard is present that supports the use of 3-D audio, then DirectX will use spatial location algorithms provided by the soundcard drivers. In the absence of such hardware, DS3D will implement its own software algorithms.

Early implementations of 3-D audio in the DirectX suite (DirectX 3.0) were poor, with problems including distortion and processing speed overheads. Microsoft's DS3D has commands to specify varying locations in space for sound sources and for the listener, to apply Doppler shift to moving sounds and to control physical parameters such as the rate at which sound attenuates over distance. Since the release of DirectX 7.0 (and up to DirectX 9.0) Microsoft has implemented a new software 3-D sound engine with three modes, which include stereo panning, HRTF light and HRTF full. As the names suggest the first is just stereo panning (audio imaging from left to right) while the other two modes are designed to be positional sound modes using proprietary 'HRTF' filter algorithms at different levels of fidelity. Unfortunately the term 'HRTF' used in the DirectX modules only refers to what is essentially one part of a real HRTF in that interaural-time (ITD) and interaural-intensity (IID) differences for a generic model are used. The fine detail filtering (e.g. the spectral changes caused by the pinna, head and torso) that is necessary for accurate 3-D audio is not part of DS3D's 'HRTF' algorithms. DS3D is regarded as a basic positional API that relies upon other technologies (both software and hardware based) such as A3D (See Section 7.1.5), EAX (See Section 7.1.6), 3DPA (See Section 7.1.7) and Qsound (See Section 7.2.2) to supplement it for efficient and accurate 3-D positional effects computation.

7.1.3 Open Audio Library

The Open Audio Library (OpenAL) is an initiative designed to provide a cross-platform open source API solution for programming 2D and 3D audio. As the name implies, OpenAL is analogous in many ways to SGI's OpenGL, a standard for specifying high-quality 3-D graphics. As a general audio API, OpenAL is an alternative to DirectSound3D and owes its origins to the discontent developers had with the early implementations of Microsoft's audio API. As it stands however, OpenAL 1.8 only supports some of the recommendations put forth in the IA-SIG 3-D audio level 1 guidelines. IA-SIG supported features are distance, listener orientation, Doppler and sound radiation. This API is yet to fulfil the requirements for 3-D audio outlined by the IA-SIG, but its acceptance among some of the leading software and hardware companies (e.g. Creative, Epic Games and LucasArts) and its rapid development growth highlights the need to remain aware of this audio API.

7.1.4 Miles Sound System (RSX 3D)

Miles Sound System provides a unified software API for audio in PC games, incorporating support for all versions of EAX, A3D and D3D. It also incorporates its own software 3-D audio solution, RSX 3D, which is comparable to earlier versions of D3D but with more efficient algorithms. As it stands, this software based 3-D audio solution is limited to surround sound reproduction (6 sources in a horizontal arrangement) over headphones or any two speaker array.

7.1.5 Aureal A3D

Aureal's A3D relies upon both a software API and dedicated hardware to create 3-D positional audio. The hallmark of this product was its early implementation of non-individualized HRTF processing and wave-tracing technology designed to create an immersive environment for the PC entertainment market. The sound wave-tracing used was akin to ray-tracing light in 3-D graphics rendering in order to recreate photo-realistic scenes. However, this technology had high computational overheads. Although a promising product, Aureal's unfortunate demise saw an end to any further development with this technology. Creative Labs and Nvidia have taken over the intellectual property rights from Aureal, but no announcements have been made regarding any further developments with the technology.

7.1.6 Creative Labs EAX

EAX stands for Environmental Audio Extension, and as its name implies aims to synthetically recreate the aural experience of being in a real environment. Like Aureal's A3D (see Section 7.1.5), Creative Labs' EAX is both a software API and dedicated hardware solution for providing an immersive entertainment environment for the PC games industry. However, EAX is an extension to DS3D and relies upon it for true positional 3-D audio. As it stands, EAX (rev 3.0) only provides the listener with environmental filtering effects to give the impression that the listener is for example in a cave, under water or in a church cathedral.

7.1.7 Sensaura 3DPA

Sensaura licenses its technology to the major audio chip manufacturers (i.e. for PC soundcards). Sensaura's approach to producing 3D audio technology for headphones (Virtual Ear™) appears to be based on a strong theoretical and research background. They have addressed the issues posed by the use of non-individualized HRTFs through the use of scaling techniques.

Sensaura's approach to creating high fidelity 3D audio is to provide listeners with a baseline HRTF library set that is representative of the population. The Virtual Ear™ product then allows for the altering of HRTF characteristics to suit differences in individual human physical dimensions. Sensaura has chosen four physical features that contribute to the shaping of auditory cues that include Ear size, Head size, Concha depth and Concha shape.

The external ear has been modelled as a series of directionally-dependant resonators with the sum of these effects being represented in the HRTF. As ear size is changed, the relative contribution of these resonators is also altered accordingly, which results in spectral scaling in the frequency domain (See Figure 15). Changing the ear size to suit the users results in either compression or expansion of the original frequency response. This scaling is done using linear interpolation.

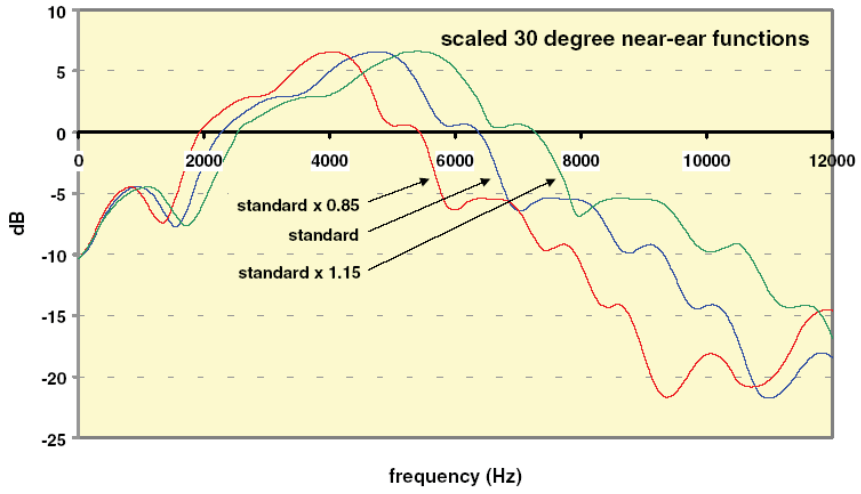


Figure 15: Spectral scaling showing examples of expansion (standard X 1.15) and compression (standard X 0.85) which result in changes to the spectral shape consistent with changes that occur as a result of differences in ear size (reproduced from Sensaura, 1999).

The effect of head size is concentrated mostly on the duration of the Interaural Time Difference (ITD) (e.g. the larger the head, the longer the ITD). However, the head also acts as an attenuator, decreasing the intensity of high frequencies that arrive at the far ear.

Sensaura's technology uses mathematical algorithms to model the ITD with differing head size. The function plotted in Figure 16 shows ITD values as a function of an increase and decrease in azimuth angle. This function is linearly scaled according to head size. If the head size is smaller, the function is compressed. The function is expanded for larger head sizes.

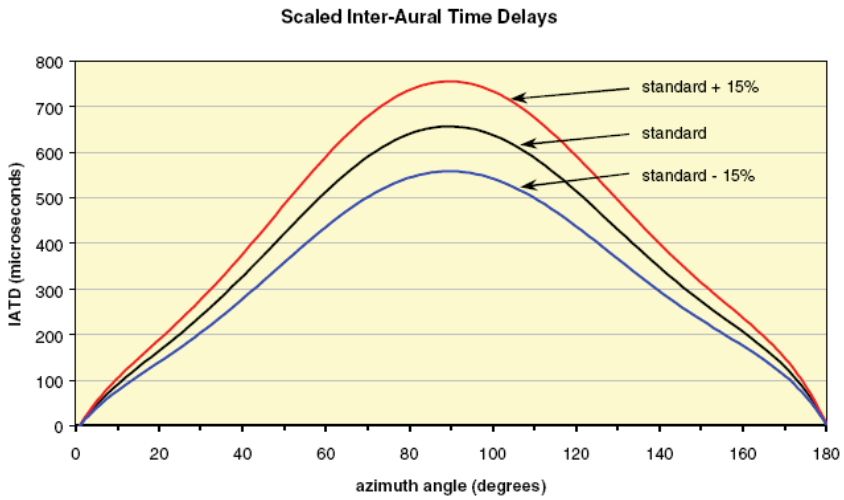


Figure 16: ITD scaling to accommodate changes in head size showing examples of expansion (standard +15%) replicating an increase in head size and compression (standard –15%) replicating a decrease in head size (reproduced from Sensaura, 1999).

Most people's ears consist of the same collection of physiological features, but the relative size of each of these individual features can vary between individuals. The concha (see Figure 8) is one of the primary resonant cavities of the external ear and can vary between individuals in terms of depth. Some individuals can have a relatively shallow concha, whilst others can exhibit a more substantial cavity with greater depth (see Figure 17). As a result of the acoustic characteristics of the concha, largely influenced by its depth, HRTFs generally have a peak amplitude around 5 kHz. When the concha has a short length, the peak in the HRTF has been found to move higher in frequency. To model this tendency, Sensaura use a simple displacement method to move the peak of the HRTF to a more suitable frequency for the concha depth determined.

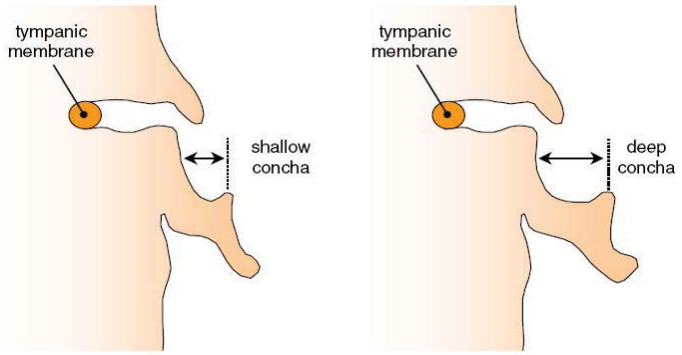


Figure 17: Plan section of the external ear (pinna) showing differing depths of the concha, with a shallow depth concha on the left and a deeper concha cavity of the right (reproduced from Sensaura, 1999)

The shape of the concha also changes across listeners. Sensaura have stated that some concha have a rim present that increases the primary resonances (see Figure 18). The depth of the concha determines the fundamental frequency so this must remain unaltered. The effect of increased primary resonances (caused by the presence or absence of a rim) is achieved by linear scaling of the logarithmic amplitude of the frequency response of the HRTF (See Figure 19).

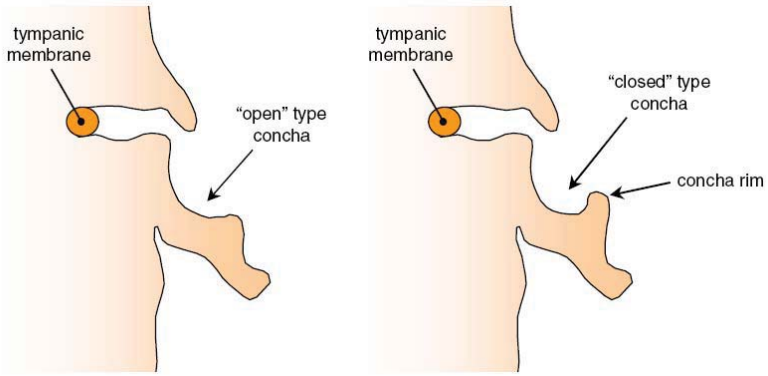


Figure 18: Plan section of the external ear (pinna) showing differing shapes of the concha, with an open type concha on the left and a closed type concha cavity with the presence of a concha rim of the right (reproduced from Sensaura, 1999)

Although scaling is beneficial to accommodate for idiosyncrasies in auditory cues, it does have some limitations. One problem is that scaling of the HRTF is performed on a reference non-individualized HRTF. The scaling procedures can only alter and realign the information already present in the reference HRTF. Such techniques cannot provide any additional information required by the listener that may not be present in the generic HRTF. The assumption is therefore that most of the critical features are present in every

HRTF and that the alignment is what varies between individual listeners. Also, these techniques tend to have an impact over most of the frequency range. It is possible that whilst some features may be realigned and result in a better match, other features may be shifted to frequency regions that result in a mismatch.

These techniques described are methods for improving the match of HRTFs to each listener without the need to measure the individual listeners HRTFs. Whilst the approach of Sensaura to include an ability to scale the various components of the HRTFs has a robust theoretical basis, there are conflicting reports as to the success of the approach, which is often described as falling short of the expectations generated based on its theoretical underpinnings.

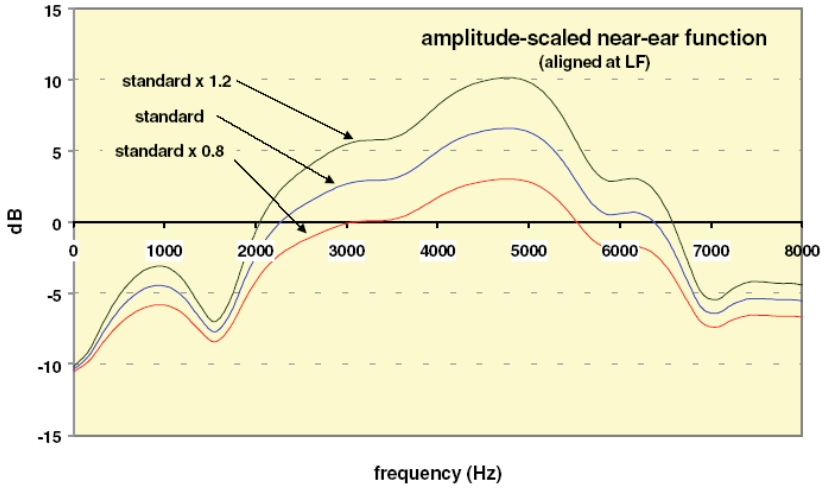


Figure 19: Amplitude scaling to accommodate changes in concha shape showing examples of gain (standard X 1.2 and standard X 0.8) replicating changes in concha shape, primarily the presence or absence of a concha rim (reproduced from Sensaura, 1999)

7.1.8 SLAB

SLAB is a non-commercial software API in development at the Spatial Auditory Displays Laboratory at NASA Ames Research Centre. Specifically designed for research applications, SLAB's strengths are that it can accept individualized HRTFs (it also comes with a set of generic HRTFs), operates at near real-time speed and has provision for direct input from a head-tracking device. The software is only applicable for playback over headphones (the preferred delivery for 3-D audio). The facility to use individualized HRTFs also allows appropriate headphone calibration corrections to be incorporated into the digital filters thus providing maximum spatial fidelity. As of the current iteration of the software (v. 5.2.1) a time delay is still present, around 24ms. However this delay can be longer (up to 42ms) dependent on the efficiency of the soundcard device drivers and software operating system used. Provision for input from a head-tracking device, also allows dynamic updating of the listener's head-to-sound-source orientation. A certain advantage of SLAB is that the software API allows for server operation. In this mode of

operation the software rendering can be completed on a dedicated PC freeing up the CPU resources on the host machine for other processes (e.g. head-tracking, final stage sound output). As this product is not directed toward the entertainment or PC games industries, it does not have provisions for special effects processing such as object obstruction/reflections and Doppler-shift effects.

7.1.9 AM3D Diesel Power Engine

AM3D's Diesel Power is a commercially available software API and engine (pre-compiled rendering software program) produced primarily for the PC games and entertainment market and conforms to IA-SIG requirements (see Section 7.1.1). It uses non-individualized HRTFs designed in collaboration with the Acoustics Laboratory at Aalborg University. The HRTF database has been developed from two sources. One source is real-person recordings collected on a single individual and the other is a higher resolution database collected using an artificial head developed at Aalborg University (see Section 3.4.2). The engine is designed to operate in real-time (at 22 or 44.1 kHz playback sampling rates). The intended output medium is selectable from either headphones, or multiple speaker arrays (2 or 4 using cross-talk cancellation). Provision is made for headphone calibration adjustment through a list of common headphone models. In addition to providing true positional 3-D audio for sound sources, Diesel Power also provides computations for object occlusion/reflection, Doppler-shift and source sound-field radiation. The API also gives provision for direct input from head-tracking devices. AM3D's 3-D audio API is also used in other commercial products (e.g. as part of an Electronic Warfare system for military aircraft manufactured by TERMA).

7.2 Multi-Channel, Surround Sound Spatial Audio Technologies

7.2.1 Dolby, DTS, Dolby Headphone and SRS Labs

The entertainment industry has been utilising 'spatial' audio technologies for some time. These technologies are based on multiple speaker set-ups and the design aim is to create an immersive audio environment to supplement video or film footage and not to provide true 3-D positional audio. Dolby Digital and Digital Theatre Sound (DTS) are six-channel digital surround sound systems which both use the 5.1 speaker format. The format consists of three speakers across the front and two speakers in the rear. The .1 is a sixth channel which produces low frequency effects that are sent to a subwoofer. Dolby Digital uses the AC-3 file format, which any Dolby Digital Decoder can decode to produce 5.1 audio. Dolby Digital is the technical name for Dolby's multi-channel digital sound coding technique, more commonly referred to as Dolby 5.1. DTS is also a multi-channel recording format (5.1 channels also) used in the entertainment industry, although the algorithm it uses produces higher quality sound than AC-3 (less compression) but at the expense of being less efficient. Dolby Headphone technology provides 3-D audio but only by reproducing 6 channels over headphones in order to simulate a real 5.1 speaker set-up in a horizontal plane arrangement. SRS Lab's technology also focuses on multi-channel (5.1 or 6.1 channels) audio playback but through reproducing multi-channel playback from mono or stereo sources (e.g. stereo expansion).

7.2.2 Qsound Q3D

Qsound is a company that license their 3-D audio technology to other 3rd-party companies that manufacture audio products (e.g. soundcards). Their technology comes in two streams, Q1 and Q2. Q1 aims to reproduce 3-D audio through two standard speakers using proprietary algorithms that do away with the need for cross-talk cancellation to generate 3-D audio in freestanding speakers. Q2 technology is Qsound's solution for providing 3-D audio over headphones. This technology is based on HRTF measurements taken from dummy-head recordings and implemented with computationally efficient algorithms. Qsound's market is predominately PC entertainment and as such their emphasis is on efficiency over precision.

7.3 Dedicated Hardware, True Positional 3-D Audio Solutions

7.3.1 DSP Microprocessors

At the heart of 3-D audio technology are the mathematical calculations. How these calculations are processed is one of the key distinctions among the various products and technologies currently available. Dedicated 3-D audio hardware is the exclusive domain of the high-end systems (e.g. Lake Huron, AuSIM3D and TDT), with the other end of the spectrum being predominately software based. Software solutions are where the mathematical calculations are handled by a computer's microprocessor or CPU (central processing unit), with only some tasks routed to any available DSP chips on board a soundcard. The filtering involved in positional 3-D audio processing (i.e. convolution) is out of the domain of the simple DSP capabilities on most soundcards and is therefore left for the CPU to process. The common CPU (e.g. Intel or AMD) is designed to run business and other general applications or data manipulation operations. They are not optimised for algorithms or mathematical operations (although they can process both but not necessarily efficiently). The *SHARC®* range of DSP chips from Analog Devices are microprocessors specifically designed to handle digital signal processing tasks, or complex mathematical operations like convolution operations. These microprocessors are not specifically designed for 3-D audio per se, but are dedicated to handling the required computations that are necessary for the filtering operations involved in true positional 3-D audio. At the simplest level, DSP microprocessors are capable of conducting the core mathematical steps involved in the convolution operation. This gives the dedicated DSP microprocessor a computational advantage beyond the comparable processor clock operating speed specifications of most of today's CPUs. It should be understood however, that DSP microprocessors like the *SHARC®* range, are far from a complete 3-D audio solution. They are merely a key component in a real-time positional 3-D audio system. An examination of the components of the high-end dedicated 3-D audio hardware systems will most often list SHARC DSPs (or equivalent DSP microprocessors from companies such as Texas Instruments).

7.3.2 Tucker Davis Technologies

Tucker Davis Technologies (TDT) manufacture three digital-signal-processing (DSP) devices that are capable of generating individualized 3D audio: the PD1 Power Dac

Convolver, the RP2.1 Enhanced Real-time Processor, and the RL2 Stingray Pocket Processor.

The RP2 Real Time Signal Processor, whilst not designed specifically for 3D audio, does have certain capabilities optimised for 3D audio production. It allows the user to load an entire HRTF filter bank into its memory and plays appropriate location sound sources where only the azimuth and elevation need to be specified. It also has interpolation algorithms that provide movement of the sound source in virtual acoustic space. TDT has also released its Stingray Pocket Processor that is part of the RP2 group of products. It provides a scaled down version of the functions as the RP2 Real Time Signal Processor but is portable and battery operated.

Another product developed by TDT is the PD1 Power SDAC (see Figure 20) that is specially targeted for 3D audio production. It provides the user with Windows based software that can be used to program the hardware to generate and playback 3D audio to the user's specifications. The modular design of the TDT systems allows for the addition of extra DSP hardware required to meet the needs of 3D audio production. For example the PD1 Power SDAC can be partnered with the PD1 Power SDAC Convolver to provide a platform for 3D audio applications. The purpose of this sub unit is to provide extra speed during calculations involving real-time 3D audio generation, variable sample rates up to 175 kHz, head tracking for head movement compensation and a completely programmable capability.

The PD1 Power Dac Convolver can be fitted with up to 28 convolver blocks, each capable of convolving an audio signal with an FIR filter of up to about 400 taps (or 200 taps in stereo mode) at a sampling rate of 50 kHz in real time. Each PD1 can be fitted with up to four DAC outputs, which could provide left- and right-ear stimulation for two listeners. If 128-tap head-related transfer functions (HRTFs) were used, a single PD1 could simultaneously generate up to 14 separate 3D sound sources for each of two listeners. Up to four sound signals could be externally generated and fed into the PD1 via ADC inputs.

Figure 20: Tucker Davis Technologies – PD1 Power SDAC Convolver

The PD1 Power Dac Convolver interfaces with TDT's AP2 Array Processor Card, which is PC hosted. The AP2 can be used to generate audio signals and to store and manipulate HRTFs. Both the PD1 and the AP2 are programmable in C or Pascal using libraries provided by TDT. TDT also provide a suite of Windows-based support and development programs that can be used to facilitate PD1 and AP2 programming for general DSP or 3D audio applications.

The RP2.1 Enhanced Real-time Processor (see Figure 21) is a more recent TDT product. It is capable of generating a single 3D sound source for one listener (128-tap HRTFs running at 50 kHz). The RP2.1 has built-in signal generation capabilities and can store a set of HRTFs up to about 32 MB in size. It can be programmed using a range of TDT software products via a PC interface.

Figure 21: Tucker Davis Technologies - RP2.1 Enhanced Real-time Processor

7.3.3 DSTO 3D Audio and TDT Hardware

The fidelity of a 3D audio system is critically dependent on the quality of the individualized HRTFs with which it operates. DSTO has directed considerable effort toward the development of HRTF measurement and processing techniques that allow the efficient production of high-fidelity 3D audio. DSTO's HRTFs are measured using a custom-built stimulus-delivery system and TDT signal generation and recording equipment (the PD1 Power Dac Convolver). These HRTFs are further processed to achieve effects such as headphone compensation using DSTO-developed software.

DSTO's 3D audio is generated using the TDT PD1 Power Dac Convolver programmed by DSTO software that makes use of TDT's C libraries. A central component of this software is its HRTF interpolation routine, which generates HRTFs for locations at which measurements have not been made and allows a smooth rendition of virtual auditory space. This routine has been developed within DSTO and operates independently of the TDT equipment.

It should therefore be noted that whilst DSTO utilises TDT hardware components the generation of 3D audio is reliant upon software developed at DSTO to produce the fidelity as reported by DSTO (see Sections 3.2, 4.1 and 5 for a review of DSTO laboratory and simulation experiments).

7.3.4 AuSIM Engineering Solutions AuSIM3D™

The AuSIM3D™ systems from AuSIM Engineering Solutions consist of software algorithms performed on dedicated hardware (e.g. AuSIM Goldminer™, see Figure 22) that can simulate the propagation of audio through a modeled environment from an emitter to a human listener. The algorithms form a hybrid of physically based models and empirically measured approximations. AuSIM3D™ builds their audio simulation from four key components: signal generation, sound source emission, wave propagation, and sound perception.

Figure 22: AuSim Goldminer system

AuSIM3D™ contains some rudimentary FM signal generators, but directly couples physically based synthesis to physically based propagation models.

Sound source emission involves modeling how sound waves propagate away from their emitter (see Figure 23). This component is important for realistic simulation of actual speaker drivers (e.g. modeling the propagation cone for a specific speaker driver). AuSIM's sound wave propagation model is in development, but their plans involve the full simulation of the propagation of sound waves through complex 3-D geometric environments. Currently, only a distance model is implemented (assumes a constant medium between source and listener).

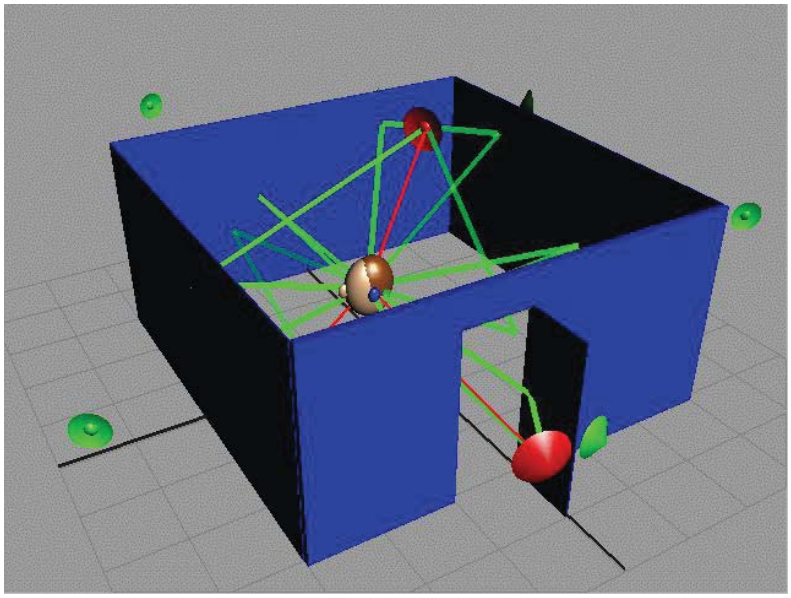


Figure 23: AuSim wave tracing simulation in a complex 3-D environment

The final component of AuSIM3D™ is the sound perception model. Fortunately, AuSIM acknowledge this stage as the most important. AuSIM's HRTF implementation is a hybrid between measured impulse responses and parametric physically based models of the human hearing system. Ausim GoldSeries Systems are integrated packages incorporating the AuSim3D™ technology. The specifics of AuSim's Goldminer™ include support for 128 (expandable to 256) sound sources, up to 4 listeners (expandable to 32) and provision for using custom filters up to 512-taps in length (high-fidelity HRTFs) permitting individualized HRTFs to be used as an alternative to the generic filters supplied with the AuSIM3D systems.

7.3.5 Lake Technology Huron 20

Lake Technology (whose origin was in Australia) was the initial developer behind Dolby Headphone technology. The basis of this research was their own Huron digital convolution workstation, which is a complete hardware and software product aimed at the virtual reality simulation and acoustic research markets. The Huron system is capable of delivering 3-D positional audio over headphones or multi-speaker arrays, with support for up to 50 speakers. Lake's Space Array and Aniscape software controls multi-speaker arrays to position virtual sound sources with trajectories and Doppler effects over wide listening areas. Binscape is the software tool that allows the position/trajectory of the 3-D audio to be specified and manipulated. Headscape is a software tool that provides the listener with head-relative 3-D audio over headphones. By tracking the listener's head orientation and providing smooth switching between binaural filters (with a 1.5 degree separation), Headscape can provide a 3-D audio image where the listener is free to move his/her head. Multiple users can be accommodated with a single system in both complex geometrically modelled and measured impulse response environments. Multiscape is the

multi-user software tool for headphone and speaker arrays that allows users to interact in a virtual audio world, also allowing users to interact with each other and the virtual environment.

The hardware capabilities of the Huron system include real-time rendering with multiple sound-source input, Doppler effects and sound source directivity at up to 96kHz sampling rate. The Huron system provides non-individualized positional audio, as well as the facility for recording custom impulse responses (geared toward room acoustic simulation) and implementing these custom filters (including individualized HRTF impulse responses). The system is capable of real-time convolution of up to 278,244 tap FIR filters (necessary for complex room acoustic simulations). The Huron workstation is a scalable-modular system that can accommodate extra DSP boards to cover the requirements from entry level DSP with a single listener to multiple-user operations with integrated head-tracking and interpolating filter responses (with head movements).

7.3.6 Vast audio

Another Australian company with some expertise in the design and construction of 3-D audio systems is VAST Audio Pty Ltd (a spin off from the Auditory Neuroscience Laboratory at the University of Sydney). VAST Audio is currently developing two patented platform technologies. First, a 3D-audio technology that allows simpler measurement of the listener's ear shape for tailoring spatial audio to the individual listener. If this technology matures it might replace the need to conduct individual recordings to obtain accurate 3-D audio. Second, a recording and encoding technology that allows a perceptually exact copy of the original 3D sound field to be reproduced for any listener in an efficient manner.

VAST Audio has expertise in the areas of:

- software and hardware design and development for the rendering of 3D audio for indoor/outdoor augmented reality and virtual reality systems and also for spatialised audio communication systems.
- circuit and chip design to manage implementations of devices to render augmented or virtual worlds or to support other communications functions.

7.3.7 Conclusions

The choice of commercial 3D audio system is primarily driven by the requirements associated with the potential application of the system. For situations where all that is required is a different audio experience over headphones than the conventional stereo currently available (e.g. games), software implementations of 3D audio provide a working solution. Where greater fidelity regarding the spatial location of the sound is required (i.e. where a sound source can be localised in a definite location in virtual acoustic space) then a more expensive hardware solution is required, possibly with individualized HRTFs.

8. Summary

The intent of this report has been to provide an overview of relevant material associated with the construction of 3-D audio systems. This report is not intended to be a detailed design guide for the construction of a 3-D audio system, since too many of the detailed design decisions are dependent on the specific application and the operational environment (i.e. platform and cockpit). Some background information describing the perceptual basis of 3-D audio is provided as a basis to help understand some of the requirements associated with the construction of a 3-D audio system. Also included is a brief review of some of the research associated with the development of 3-D audio, including a review of 3-D audio research at DSTO that provides further background to explain how some of the requirements for a 3-D audio system were derived. Following this background is a discussion of the hardware and software requirements for the implementation of 3-D audio. These requirements are in general terms and include those that should form the basis of any 3-D audio system. Finally, a brief survey of current commercial 3-D audio systems is reported, and some discussion of issues that require further consideration is included.

Some of the issues that have been raised need careful consideration during the detailed design process to ensure that an appropriate system is delivered. Like any system being fitted to an existing platform there will be some integration issues, but the intent of this report has been to provide some early indications regarding where these issues might arise. It is also important to note that whilst we possess a great deal of knowledge regarding the construction and implementation of 3-D audio systems, this technology is still relatively young and there remain some areas that would benefit from continued research. In addition to issues associated with the design and construction of the 3-D audio system there are a number of other factors that need to be considered in order to ensure successful implementation e.g. background noise, pilot workload. Like any new technology an appropriate risk reduction strategy is to maintain a continuing research program through to delivery into service to ensure the complete success of this capability and to ensure that 3-D audio lives up to the promise that has been demonstrated in the laboratory and simulation environments.

9. References

Begault, D. R. (1994). *3-D sound for virtual reality and multimedia.* Chesnut Hill, MA: Academic Press.

Begault, D. R. (1995). Virtual acoustic displays for teleconferencing: Intelligibility advantage for telephone grade audio. *Audio Engineering Society 98th Convention Preprint 4008*. New York, AES.

Begault, D. R. & Wenzel, E. M. (1993). Headphone localization of speech. *Human Factors, 35,* 361-376.

Begault, D. R. & Erbe, T. (1994). Multichannel spatial auditory display for speech communications. *Journal of the Audio Engineering Society, 42 (10)*, 819-826.

Begault, D. R. & Pittman, M. T. (1996). Three-Dimensional audio versus head-down traffic alert and collision avoidance system displays. *International Journal of Aviation Psychology, 6 (1), 79-93*

Begault, D. R., Wenzel, E. M., & Lathrop, W. B. (1997). Augmented TCAS advisories using a 3-D audio guidance system. *Proceedings of the Ninth International Symposium on Aviation Psychology,* (pp. 353-357). Columbus, OH: Ohio State University.

Blauert, J. (1969/70). Sound localization in the median plane. *Acustica, 22,* 205-213.

Bles, W. (2004). Spatial disorientation countermeasures – Advanced problems and concepts. In F. H. Previc & W. R Ercline (Eds.), *Spatial disorientation in aviation.* (pp. 509-541). Reston, Virginia: American Institute of Aeronautics and Astronautics.

Bolia, R. S. (2003). Spatial audio displays for air battle managers: Does visually cueing talker location improve speech intelligibility? *Proceedings of the 12th International Symposium on Aviation Psychology,* (pp. 136-139).

Bolia, R. S., D'Angelo, W. R., & McKinley, R. L. (1999). Aurally aided visual search in three-dimensional space. *Human Factors, 41 (4),* 664-669.

Bronkhorst, A. W. (1995). Localization of real and virtual sound sources. *Journal of the Acoustical Society of America, 98 (5)*, 2542-2553.

Bronkhorst, A. W. (1999). Adapting head-related transfer functions to individual listeners. *Journal of the Acoustical Society of America, 105 (2)*, 1036.

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acustica, 86,* 117-128.

Bronkhorst, A. W. & Plomp, R. (1992). Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing. *Journal of the Acoustical Society of America, 92*, 3132-3138.

Brungart, D. S., Ericson, M. A., & Simpson, B. D. (2002). Design considerations for improving the effectiveness of multitalker speech displays. *Proceedings of the 2002 International Conference on Auditory Display,* (pp. 2.1-2.7).

Brungart, D. S., & Simpson, B. D. (2003). Optimizing the spatial configuration of a seven-talker display. *Proceedings of the2003 International Conference on Auditory Display,* (pp. 3.1-3.4).

Burger, J. F. (1958). Front-back discrimination of the hearing system. *Acustica, 8,* 301-301.

Butler, R. A., & Planert, N. (1976). The influence of stimulus bandwidth on the localization of sound in space. *Perception & Psychophysics, 19,* 103-108.

Cheng, C. I. & Wakefield, G. H. (2001). Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space. *Journal of the Audio Engineering Society,* 49, 231-249.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America, 25*, 975-979.

Crispien, K., & Ehrenberg, T. (1995). Evaluation of the cocktail-party effect for multiple speech stimuli within a spatial audio display. *Journal of the Audio Engineering Society, 43*, 932-941.

Davis, R. J., & Stevens, S.D.G. (1974). The effect of intensity on the localization of different acoustical stimuli in the vertical plane. *Journal of Sound and Vibration, 35,* 223-229.

Drullman, R., & Bronkhorst, A. W. (2000). Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation. *Journal of the Acoustical Society of America, 107 (4)*, 2224-2235.

Ericson, M. A. & McKinley, R. L. (1997). The intelligibility of multiple talkers separated spatially in noise. In R. H. Gilkey & T. R. Anderson (Eds.), *Binaural and spatial hearing in real and virtual environments* (pp. 701-724). New Jersey USA, Erlbaum.

Flanagan, P., McAnally, K. I., Martin, R. L., Meehan, J. W., & Oldfield, S. R. (1998). Aurally and visually guided visual search in a virtual environment. *Human Factors*, *40 (3)*, 461-468.

Freedman, S. J., & Fisher, H. G. (1968). The role of the pinna in auditory localization. In S. J. Freedman (Ed.), *The neuropsychology of spatially oriented behaviour* (pp. 135-152). Homewood, Illinois: The Dorsey Press.

Gardner, M. B., & Gardner, R. S. (1973). Problem of localization in the median plane: effect of pinnae cavity occlusion. *Journal of the Acoustical Society of America, 53 (2)*, 400-408.

Good, M. D. & Gilkey, R. H. (1996). Sound localization in noise: The effect of signal-to-noise ratio. *Journal of the Acoustical Society of America, 99 (2)*, 1108-1117.

Griffiths, S., Watt, J. & Parker, S. P. A. (2005). Impact of spectral content on the design of signals for use in 3-D auditory displays. DSTO Technical Report. Melbourne, Australia: Aeronautical and Maritime Research Laboratory.

Haas, E. C. (1998). Can 3-D auditory warnings enhance helicopter cockpit safety? *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting,* (pp. 1117-1121). Santa Monica, CA: Human Factors and Ergonomics Society.

Haas, E. C., Gainer, C., Wightman, D., Couch, M. & Shilling, R. (1997). Enhancing system safety with 3-D audio displays. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting* (pp. 868-872). Santa Monica, CA: Human Factors and Ergonomics Society.

Hebrank, J., & Wright, D. (1974). Spectral cues used in the localization of sound sources in the median plane. *Journal of the Acoustical Society of America, 56, 1829-1834*.

Jin, C., Leong, P., Leung, J., Corderoy, A., & Carlile, S. (2000). Enabling individualized virtual auditory space using morphological measurements. *Proceedings of the First IEEE Pacific-Rim Conference on Multimedia (2000 International Symposium on Multimedia Information Processing),* 235-238.

Jin, C., van Schaik, A., Best, V. & Carlile, S. (2003). Perceptual spatial-audio coding. *Proceedings of the 2003 International Conference on Auditory Display,* 255-258.

King, R. B., & Oldfield, S. R. (1997). The impact of signal bandwidth on auditory localization: Implications for the design of three-dimensional audio displays. *Human Factors, 39 (2),* 287-295.

Lyons, T.J., Gillingham, K.K., Teas, D.C., Ercoline, W.R., & Oakley, C. (1990). The effects of acoustic orientation cues on instrument flight perfromance in a flight simulator. *Aviation, Space, and Environmental Medicine,* 699-706.

MacDonald, J. A., Balakrishnan, J.D., Orosz, M.D., & Karplus, W. J. (2002). Intelligibility of speech in a virtual 3-D environment. *Human Factors, 44 (2),* 272-286.

Makous, J. C., & Middlebrooks, J. C. (1990). Two-dimensional sound localization by human listeners. *Journal of the Acoustical Society of America, 87 (5),* 2188-2200.

Martin, R. L., McAnally, K. I, & Senova, M. A. (2001). Free-field equivalent localization of virtual audio. *Journal of the Audio Engineering Society, 49 (1/2)*, 14-22.

Martin, R. L., Parker, S. P. A., McAnally, K. I., & Oldfield, S. R. (1996). The abilities of listeners to localise Defence Research Agency auditory warnings. *NATO Advisory Group for Aerospace Research and Development-CP-596*, 9, 1-7.

McAnally, K. I., Bolia, R. S., Martin, R. M., Eberle, G., & Brungart, D. S. (2002). Segregation of multiple talkers in the vertical plane: Implications for the design of a multiple talker display. *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting* (pp. 588-591). Santa Monica, CA: Human Factors and Ergonomics Society.

McAnally, K. I., & Martin, R. M. (2004). *The efficiency with which front-back and elevation are resolved by head-movements.* Unpublished manuscript.

Middlebrooks, J. C. (1999a). Individual differences in external-ear transfer functions reduced by scaling in frequency. *Journal of the Acoustical Society of America, 106 (3),* 1480-1492.

Middlebrooks, J. C. (1999b). Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. *Journal of the Acoustical Society of America, 106 (3),* 1493-1510.

Middlebrooks, J. C., & Green, D. M. (1990). Observations on a principle components analysis of head-related transfer functions. *Journal of the Acoustical Society of America, 92,* 597-599.

Middlebrooks, J. C., Makous, J. C., & Green, D. M. (1989). Directional sensitivity of sound-pressure levels in the human ear canal. *Journal of the Acoustical Society of America, 86,* 89-108.

Minnaar, P., Olesen, S. K. Christensen, F. & Møller, H. (2001). Localization with binaural recordings from artificial and human heads. *Journal of the Audio Engineering Society., 49 (5),* 323-336.

Møller, H., Sorensen, M. F., Hammershoi, D., & Jensen, C. B. (1995). Head-related transfer functions of human subjects. *Journal of the Audio Engineering Society., 43 (5),* 300-321.

Møller, H., Sorensen, M. F., Jensen, C. B. & Hammershoi, D. (1996). Binaural technique: Do we need individual recordings? *Journal of the Audio Engineering Society., 44 (6),* 451-469.

Musicant, A. D., & Butler, R. A. (1984). The influence of pinnae based spectral cues on sound localization. *Journal of the Acoustical Society of America, 75 (4)*, 1195-1200.

Nelson, W. T., Bolia, R. S., Ericson, M. A., & McKinley, R. L. (1998). Monitoring simultaneous presentation of spatialized speech signals in virtual acoustic environments. *Proceedings of the 1998 IMAGE conference* (pp. 159-166).

Nelson, W. T., Bolia, R. S., Ericson, M. A., & McKinley, R. L. (1999). Spatial audio displays for speech communications: A comparison of free-field and virtual acoustic environments. *Proceedings of the Human Factors and Ergonomics Society 43rd Annual*

Meeting (pp. 1202-1206). Santa Monica, CA: Human Factors and Ergonomics Society.

Nelson, W. T., & Bolia, R. S. (2003). Evaluating the effectiveness of spatial audio displays in a simulated airborne command and control task. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 202-206). Santa Monica, CA: Human Factors and Ergonomics Society.

Noble, W., Byrne, D., & Lepage, B. (1994). Effects on sound localization of configuration and type of hearing impairment. *Journal of the Acoustical Society of America, 95 (2)*, 993-1005.

Oving, A. P., & Bronkhorst, A. W. (1999). Application of a three-dimensional auditory display for TCAS warnings. *Proceedings of the Tenth International Symposium on Aviation Psychology,* (pp. 26-31). Columbus, OH: Ohio State University.

Parker, S. P. A., Smith, S. E., Stephan, K. L., Martin, R. L., & McAnally, K. I. (2004). Effects of supplementing Head-Down Displays with 3-D audio during visual target acquisition. *International Journal of Aviation Psychology, 14 (3),* 277-295.

Patterson, R.D. & Ditta, A. J. (1999). Extending the domain of auditory warning sounds: Creative use of high frequencies and temporal asymmetry. In N. A. Stanton & J. Edworthy (Eds.), *Human factors in auditory warnings* (pp. 73-88). Hants, UK: Ashgate.

Perrot, D. R., Cisneros, J., McKinley, R. L., & D'Angelo, W. R. (1995). Aurally aided detection and identification of visual targets. *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting,* (pp. 104-108). Santa Monica, CA: Human Factors and Ergonomics Society.

Pralong, D., & Carlile, S. (1994). Measuring the human head-related transfer functions: A novel method for the construction and calibration of a miniature "in-ear" recording system. *Journal of the Acoustical Society of America, 95 (6),* 3435-3444.

Ricard, G. L., & Meirs, S. L. (1994). Intelligibility and localization of speech from virtual directions. *Human Factors, 36 (1),* 120-128.

Sensaura, Inc. (1999). *Virtual Ear Technology.* Retrieved October 2, 2006, from http://www.sensaura.com

Sextant. (1997). *AUDIS multipurpose AUditory DISplay for 3-D hearing applications, Second progress report.*

Shigeno, S., & Oyama, T. (1983). Localization of speech and non-speech sounds. *Japanese Psychological Research, 25*, 112-117.

Stanton, N. A., & Edworthy, J. (1999). Auditory warnings and Displays: An overview. In N. Stanton & J. Edworthy (Eds.) *Human factors in auditory warnings.* (pp. 3-30). Aldershot: Ashgate.

Stephan, K. L., Smith, S. E., Parker, S. P. A., Martin, R. L., & McAnally, K. I. (2003). Auditory warnings in the cockpit: An evaluation of potential sound types. In G. Edkins & P. Pfister (Eds.), *Innovation and consolidation in aviation.* (pp. 231-241). Hants, UK: Ashgate.

Stevens, C., Brennan, D., & Parker, S. (2004). Proceedings of the 18th International Congress on Acoustics, (pp. 1821-1824). Kyoto, Japan, April.

Wenzel, E. M., Arruda, M., Kistler, D. J., & Wightman, F. L. (1993). Localization using non-individualized head-related transfer functions. *Journal of the Acoustical Society of America, 94(1),* 111-123.

Wightman, F. L., & Kistler, D. J. (1989a). Headphone simulation of free-field listening. I: Stimulus synthesis. *Journal of the Acoustical Society of America, 85(2)*, 858-867.

Wightman, F. L. (1989b). Headphone simulation of free-field listening. II: Psychophysical validation. *Journal of the Acoustical Society of America*, *85 (2),* 868-878.

Yost, W. A. (1992). The cocktail party problem: forty years later. In R. H. Gilkey & T. R. Anderson (Eds.), *Binaural and spatial hearing in real and virtual environments* (pp. 329-347). New Jersey USA, Erlbaum.

Yost, W. A., Dye, R. H., & Sheft, S. (1997). A simulated cocktail party with up to three sound sources. *Perception & Psychophysics, 58,* 1026-1036.

10. Contact details for DSTO staff

Russell Martin

Air Operations Division, DSTO
506 Lorimer St, Fishermens Bend, VIC 3207
Phone 03 9626 7115
Fax 03 9626 7084
Email Russell.Martin@dsto.defence.gov.au

Ken McAnally

Air Operations Division, DSTO
506 Lorimer St, Fishermens Bend, VIC 3207
Phone 03 9626 7251
Fax 03 9626 7084
Email Ken.McAnally@dsto.defence.gov.au

Simon Parker

Air Operations Division, DSTO
506 Lorimer St, Fishermens Bend, VIC 3207
Phone 03 9626 7269
Fax 03 9626 7084
Email Simon.Parker@dsto.defence.gov.au

Geoffrey Eberle

Air Operations Division, DSTO
506 Lorimer St, Fishermens Bend, VIC 3207
Phone 03 9626 7337
Fax 03 9626 7084
Email Geoff.Eberle@dsto.defence.gov.au

Appendix A: DSTO auditory research publications

Journal papers

Flanagan, P., McAnally, K.I., Martin, R.L., Meehan, J.W. and Oldfield, S.R. (1998) Aurally and visually guided visual search in a virtual environment. *Human Factors* 40, 461-468.

Eberle, G., McAnally, K.I., Martin, R.L. and Flanagan, P. (2000) Localization of amplitude-modulated high-frequency noise. *Journal of the Acoustical Society of America* 107, 3568-3571.

Watson, D., Martin, R., McAnally, K.I., Smith, S., and Emonson, D. (2000) Effect of normobaric hypoxia on auditory sensitivity. *Aviation, Space and Environmental Medicine* 71, 791-797.

Martin, R.L., Watson, D. Smith, S., McAnally, K.I. and Emonson, D. (2000) Effect of normobaric hypoxia on sound localization. *Aviation, Space and Environmental Medicine* 71, 991-995.

Martin, R.L., McAnally, K.I., and Senova, M.A. (2001) Free-field equivalent localization of virtual audio, *Journal of the Audio Engineering Society* 49, 14-22.

Senova, M.A., McAnally, K.I. and Martin, R.L. (2002) Localization of virtual sound as a function of head related impulse response duration. *Journal of the Audio Engineering Society* 50, 57-66.

McAnally, K.I. and Martin, R.L. (2002) Variability in the headphone-to-ear-canal transfer function. *Journal of the Audio Engineering Society* 50, 263-266.

McAnally, K.I., Watson, D.B., Martin, R.L. and Singh, B. (2003) Effect of hypobaric hypoxia on auditory sensitivity. *Aviation, Space and Environmental Medicine* 74, 1251-1255.

Martin, R.L., Paterson, M. and McAnally, K.I. (2004). Utility of monaural spectral cues enhanced by knowledge of sound-source lateral angle. *Journal of the Association for Research in Otolaryngology*, 5, 80-89.

Parker, S.P.A., Smith, S.E., Stephan, K.L., Martin, R.L. and McAnally, K.I. (2004) Effects of supplementing Head-Down Displays with 3-D audio during visual target acquisition. *International Journal of Aviation Psychology, 14(3),* 277-295.

Sandor PMB, McAnally KI, Pellieux L & Martin RL (2005) Localization of virtual sound at 4 Gz, *Aviation, Space and Environmental Medicine*, 76, 103-107.

Eramudugolla R, Irvine DRF, McAnally KI, Martin RL & Mattingley JB (2005) Directed attention eliminates 'change deafness' in complex auditory scenes. *Current Biology,* 15, 1108-1113.

Stephan KL, Smith SE, Martin RL, Parker SPA & McAnally KI (2006) Learning and retention of direct, indirect-nonarbitrary and indirect-arbitrary associations between auditory icons and denotative referents, *Human Factors*, 48, 288-299.

Carlile S, Martin RL & McAnally KI (2005) Spectral information in sound localization. International Review of Neurobiology 70, 399-434.

McAnally KI & Martin RL Spatial audio displays improve the detection of target messages in a continuous monitoring task, *Human Factors,* in press.

Eramudugolla R, Irvine DRF, McAnally KI, Martin RL & Mattingley JB Dissociation between detection and identification of change in complex auditory scenes, submitted to *Perception and Psychophysics*.

McAnally KI & Martin RL Sound localisation during illusory self-rotation, submitted to *Experimental Brain Research*.

Conference papers

Parker, S.P.A., Oldfield, S.R. and Martin, R. (1996). Design issues for aircraft auditory displays. *2nd International Symposium on Advanced an Aerospace Science and Technology.*

Martin, R.L., Parker, S.P.A., McAnally, K.I. and Oldfield, S.R. (1996) The abilities of listeners to localise Defence Research Agency auditory warnings. *NATO Advisory Group for Aerospace Research and Development-CP-596*, 9, 1-7.

Oldfield, S., Martin, R., and Parker, S. (1996) Auditory issues in simulation. *Proceedings of* S*imTecT 96,* p.73-78.

Leung, Y.K., Smith, S., Parker, S., and Martin, R. (1997) Learning and retention of auditory warnings. *Proceedings of the Fourth International Conference on Auditory Display,* p.129-133.

Flanagan, P., McAnally, K., Martin. R., Meehan, J. and Oldfield, S. (1997) Aurally and visually aided visual search in a virtual environment. *Proceedings of SimTecT 97*, 233-237.

McAnally, K.I., Martin, R.L. and Senova, M.A. (2000) The spatial fidelity of virtual audio. *Proceedings of the IEEE Pacific Rim Conference on Multimedia 2000*. (invited paper)

Stephan, K.L., Smith, S.E., Parker, S., Martin, R and McAnally, K. (2000) Auditory warnings in the cockpit: an evaluation of potential sound types. *Australian Aviation Psychology Association Symposium*.

Griffiths, S., Flanagan, P., McAnally, K. & Martin, R. (2000) Interaction between performance on aurally and visually cued visual search and continuous performance tasks. *Proceedings of the Australian Cognitive Science Conference*, Melbourne.

Senova, M., McAnally, K. and Martin, R. (2002) A psychophysical investigation of the frequency-warping coefficient. *Audio Engineering Society 22nd International Conference on Virtual, Synthetic and Entertainment Audio*. Abstract: *J. Audio Eng. Soc.* 50, 298.

McAnally, K.I., Bolia, R.S., Martin, R.M., Eberle, G. and Brungart, D.S. (2002) Segregation of multiple talkers in the vertical plane: Implications for the design of a multiple talker display. Human Factors and Ergonomics Society.

McAnally, K.I., Martin, R.L., Bolia, R.S. and Brungart, D.S. (2003) The use of virtual audio for the spatial segregation of competing speech. Eighth Western Pacific Acoustics Conference, Melbourne.

McAnally KI, Martin RL, Doman J, Eberle G & Parker SP (2005) Detection and identification of simultaneous communications in a simulated flying task. NATO RTO HFM symposium on New Directions for Improving Audio Effectiveness. RTO-HFM-123, P31-1-6.

Martin, R.L., McAnally, K.I., and Watt, J.P. (2004) Localisation of multiple sounds as a function of the duration of the inter-sound gap. *Proceeding of the 10th International Conference on Auditory Display.*

Conference abstracts

Flanagan, P., McAnally, K.I., Martin. R., Meehan, J. and Oldfield, S. (1997) Aurally and visually guided visual search. *Proceedings of the Human Factors Special Interest Group*, Department of Defence, Australia.

Flanagan, P., McAnally, K.I., Martin. R., Meehan, J. and Oldfield, S. (1997) An investigation of aurally guided and visually guided visual search in a virtual environment. *Australian Journal of Psychology* 49 S9.

Martin, R.L., Parker, S.P.A., McAnally, K.I. and Oldfield, S.R. (1997) The abilities of listeners to localise defence research agency auditory warnings. *Australian Journal of Psychology* 49 S16.

Watson, D., Martin, R., McAnally, K., Smith, S. and Emonson, D. (1998) Hearing threshold and 3D sound localization performance during exposure to simulated altitude. *Proceedings of the Human Factors Special Interest Group*, Department of Defence, Australia.

Smith, S., Parker, S., Martin, R., McAnally, K. and Stephan, K. (1999) Evaluation of sound types for use as auditory warnings. *Australian Journal of Psychology* 51 S42.

McAnally, K.I., Martin, R.L., Watt, T. and Flanagan, P. (2000) The effect of spectral roughness on sound localization. *Proceedings of the Australian Neuroscience Society* 11, 145.

Eberle, G., McAnally, K.I., Martin, R.L. and Flanagan, P. (2000) Localization of amplitude-modulated band-pass noise. *Proceedings of the Australian Neuroscience Society* 11, 142.

Senova, M.A., Martin, R.L. and McAnally, K.I. (2000) The effect of spectral detail on the localization of virtual sound. *Proceedings of the Australian Neuroscience Society* 11, 218.

Martin, R.L., Watson, D.B., McAnally, K.I., Smith, S.E. and Emonson, D.L. (2000) Effect of normobaric hypoxia on hearing. *Proceedings of the Australian Neuroscience Society* 11, 142.

Senova, M.A., Martin, R.L. and McAnally, K.I. (2000) The effect of spectral detail on the localization of virtual sound. *Australian Journal of Psychology* 52 S49.

Eberle, G., McAnally, K.I., Martin, R.L. and Flanagan, P. (2000) Localization of amplitude-modulated high-frequency noise. *Australian Journal of Psychology* 52 S32.

Paterson, M., Martin, R.L and McAnally, K.I. (2001) Utility of monaural spectral cues in binaural localization. *Proceedings of the Association for Research in Otolaryngology*.

Martin, R., Paterson, M. and McAnally, K. (2001) Correct interpretation of monaural spectral cues requires knowledge sound-source lateral angle. *Australian Journal of Psychology* 53 S59.

Smith, S.E., Stephan, K.L., Parker, S., Martin, R. and McAnally, K. (2001) Does signal-referent association strength effect learning of auditory icons? *Australian Journal of Psychology* 53 S65.

Eberle, G., McAnally, K.I., Martin, R.L. and Flanagan, P. (2001) The nature of auditory localization errors. *Australian Journal of Psychology* 53 S51.

McAnally, K.I., Watson, D.B. and Martin, R.L. (2001) Hearing sensitivity in hypobaric hypoxia. DERA symposium *Environmental variables and system design*. Farnborough, UK.

McAnally, K.I. and Martin, R.L. (2001) DSTO research in audio displays. TTCP symposium *Spatial audio displays for military aviation*, Edinburgh, UK.

Eberle, G.E., McAnally, K.I. and Martin, R.L. (2002) Do up/down reversals really exist? *Proceedings of the Association for Research in Otolaryngology*.

McAnally, K.I., Bolia, R.S., Martin, R.L., Eberle, G. and Brungart, D.S. (2002) A cocktail party in the median plane; the role of ITDs. *NATO Advanced Study Institute on Dynamics of Speech Production and Perception,* Il Ciocco, Italy.

Johnston, M. and McAnally, K.I. (2002) The influence of syntax and semantics on the intelligibility of competing speech signals. 29th Australasian Experimental Psychology Conference.

Johnston M. and McAnally, K. (2002) The influence of linguistic variables on the intelligibility of competing speech signals. Architectures and Mechanisms for Language Processing AMLaP, Tenerife.

McAnally, K., Sandor, P., Pellieux, L. and Martin, R. (2003) Sound localization in hypergravity. International Workshop on Spatial and Binaural Hearing, Utrecht. NL, June 16-19.

Martin, R., Senova, M. and McAnally, K. (2003) Efficient high-fidelity spatial audio. International Workshop on Spatial and Binaural Hearing, Utrecht. NL, June 16-19.

Eramudugolla RR, Irvine DRF, Martin RL, McAnally KI & Mattingley J (2004) The role of attention in perception of complex auditory scenes: evidence from a novel change-deafness task. Proceedings of the 2004 meeting of the Cognitive Neuroscience Society.

Lapeyronnie G, McAnally K, Roumes C, Gros R, Meehan J & Leger A (2005) Integration of alternative control and display technologies in fighter HMI: automatic speech recognition and 3d sound. New directions for improving audio effectiveness, NATO RTO-HFM-123.

McAnally KI, Sandor P, Pellieux L & Martin R (2005) Sound localisation in hypergravity. *Proceedings of the 32nd Australasian Experimental Society Conference.*

Martin R & McAnally K (2005) Sound localisation during illusory self-rotation. *Proceedings of the 32nd Australasian Experimental Society Conference.*

Eramudugolla R, Irvine D, McAnally K, Martin R & Mattingley J (2005) The role of attention in perception of complex auditory scenes: evidence from a novel change-deafness task. *Proceedings of the 32nd Australasian Experimental Society Conference.*

Eramudugolla R, Irvine DRF, McAnally KI, Martin RL & Mattingley JB (2006) Dissociation between detection and identification of change in complex auditory scenes. *Proceedings of the 33rd Australasian Experimental Society Conference.*

DSTO reports

McAnally KI (2000) Analysis of cockpit warnings in HUG and non-HUG FA-18 aircraft. AOD-CR-00/16, DSTO.

McAnally KI (2004) Analysis of cockpit noise in the Hawk-127. DSTO-TR-1634.

Martin R, McAnally K, Watt T & Flanagan P (2006) The effect of spectral variation on sound localisation. DSTO-RR-0308.

Martin R & McAnally K (2006) Interpolation of head-related transfer functions. DSTO-RR, in review.

Appendix B: Weblinks for commercial 3-D audio

Microsoft DirectSound3D (DS3D)

http://www.microsoft.com/windows/directx/default.aspx?url=/windows/directx/productinfo/overview/default.htm
http://www.3dsoundsurge.com/features/articles/ds3d.html

Interactive Audio Special Interest Group (IA-SIG) 3D Audio Workgroup

http://www.iasig.org/

Open Audio Library – OpenAL

http://www.openal.org/

Miles Sound System (RSX 3D)

http://www.radgametools.com/miles.htm

Creative Labs EAX

http://www.3dsoundsurge.com/features/articles/EAX.html

Sensaura 3DPA

http://www.3dsoundsurge.com/features/articles/Sensaura

SLAB API Spatial Auditory Displays Lab

http://human-factors.arc.nasa.gov/SLAB/

AM3D Diesel Power Engine

http://www.am3d.com/

SRS, Dolby, DTS, & Dolby Headphone

http://www.srslabs.com/
http://www.dolby.com/
http://www.dvdtech.com.au/info/technical/what_is_DTS.htm

Qsound Q3D

http://www.qsound.com/2002/

DSP Microprocessors

http://www.analog.com/index.html
http://www.bittware.com/
http://www.dspguide.com/

Tucker Davis Technologies

http://www.tdt.com/

AuSIM Engineering Solutions AuSIM3D™ AuSIM Engineering Solutions

http://www.ausim3d.com/

Vast Audio

http://www.vastaudio.com/

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION
DOCUMENT CONTROL DATA

1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)

2. TITLE

Construction of 3-D Audio Systems: Background, Research and General Requirements

3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION)

Document	(U)
Title	(U)
Abstract	(U)

4. AUTHOR(S)

Simon P.A. Parker, Geoffery Eberle, Russell L. Martin, and Ken I. McAnally

5. CORPORATE AUTHOR

DSTO Defence Science and Technology Organisation
506 Lorimer St
Fishermans Bend Victoria 3207 Australia

6a. DSTO NUMBER	6b. AR NUMBER	6c. TYPE OF REPORT	7. DOCUMENT DATE
DSTO-TR-2184	AR-014-278	Technical Report	October 2008

8. FILE NUMBER	9. TASK NUMBER	10. TASK SPONSOR	11. NO. OF PAGES	12. NO. OF REFERENCES
2005/1007413	07/225	DMO	72	71

13. URL on the World Wide Web

http://www.dsto.defence.gov.au/corporate/reports/DSTO-TR-2184.pdf

14. RELEASE AUTHORITY

Chief, Air Operations Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for public release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111

16. DELIBERATE ANNOUNCEMENT

No Limitations

17. CITATION IN OTHER DOCUMENTS Yes

18. DSTO RESEARCH LIBRARY THESAURUS http://web-vic.dsto.defence.gov.au/workareas/library/resources/dsto_thesaurus.shtml

3D Displays, Sound Localisation, Auditory Localisation, Auditory Displays, Human Machine Interface

19. ABSTRACT

Over the last few years one of the most promising advances for Human Machine Interfaces (HMI) has been the development of 3-Dimensional Audio (3-D Audio). The Air Operations Division of DSTO has been engaged in an extensive research program developing 3-D audio for the military aviation environment. This document is intended to provide some general background and to list some of the broader requirements that need to be considered when designing any 3-D audio system. Included in this report is some background information describing the perceptual basis of 3-D audio, a brief review of some of the research associated with the development of 3-D audio, a discussion of the hardware and software requirements for the implementation of 3-D audio, a brief survey of current commercial 3-D audio systems, and some discussion of the issues that require further consideration.